

Five mechanisms of sound symbolic association

David M. Sidhu¹ · Penny M. Pexman¹

Published online: 24 August 2017

© The Author(s) 2017. This article is an open access publication

Abstract Sound symbolism refers to an association between phonemes and stimuli containing particular perceptual and/or semantic elements (e.g., objects of a certain size or shape). Some of the best-known examples include the mil/mal effect (Sapir, *Journal of Experimental Psychology*, 12, 225–239, 1929) and the maluma/takete effect (Köhler, 1929). Interest in this topic has been on the rise within psychology, and studies have demonstrated that sound symbolic effects are relevant for many facets of cognition, including language, action, memory, and categorization. Sound symbolism also provides a mechanism by which words' forms can have nonarbitrary, iconic relationships with their meanings. Although various proposals have been put forth for how phonetic features (both acoustic and articulatory) come to be associated with stimuli, there is as yet no generally agreed-upon explanation. We review five proposals: statistical co-occurrence between phonetic features and associated stimuli in the environment, a shared property among phonetic features and stimuli; neural factors; species-general, evolved associations; and patterns extracted from language. We identify a number of outstanding questions that need to be addressed on this topic and suggest next steps for the field.

Keywords Sound symbolism · Iconicity · Crossmodal correspondences · Psycholinguistics

✉ David M. Sidhu
dsidhu@ucalgary.ca

¹ University of Calgary, Calgary, AB, Canada

Sound symbolism

In this review, we use the term *sound symbolism* to refer to an association between phonemes and particular perceptual and/or semantic elements (e.g., large size, rounded contours).¹ These associations arise from some quality of the phonemes themselves (e.g., their acoustic and/or articulatory features), and not because of the words in which they appear. Thus, we exclude associations deriving from patterns of phoneme use in language (i.e., conventional sound symbolism; Hinton, Nichols, & Ohala, 1994) from our definition. We also exclude associations deriving from direct imitations of sound (i.e., imitative sound symbolism; Hinton et al., 1994).² We exclude these associations because, like Hinton et al., we think they are distinct categories and that they do not necessarily share underlying mechanisms with the phenomenon we seek to explain. As illustrated in Table 1, the definition of sound symbolism that we offer here is similar to a number of other definitions for the phenomenon that can be found in the literature.

The term *association* is somewhat difficult to characterize in this context; broadly, it refers to the sense that the phonemes in question seem related to, or to naturally go along with, stimuli possessing the associated elements or features (e.g., objects of a certain size or shape). Sound symbolic associations emerge behaviorally in reports that nonwords containing certain phonemes are especially good labels for particular targets (e.g., Maurer, Pathman, & Mondloch, 2006; Nielsen &

¹ While the term *sound symbolism* is used here at the phoneme level (i.e., involving relationships between individual phonemes and semantic elements), it has also been used at the word level (e.g., Johansson & Zlatev, 2013; Nielsen & Rendall, 2011; Tanz, 1971; Westbury, 2005). These two uses are not in opposition; sound symbolic words are those whose component phonemes have a sound symbolic relationship with their meanings.

² Note that Hinton et al. (1994) used the terms *conventional* and *imitative sound symbolism* to refer to sound symbolism at the word level.

Rendall, 2011). They may also emerge on implicit tasks, such that congruent phoneme-stimuli pairings are responded to differently than incongruent pairings (e.g., Hung, Styles, & Hsieh, 2017; Ohtake & Haryu, 2013; Westbury, 2005).

These sound symbolic associations have important implications for our understanding of language. While the arbitrariness of language has long been considered one of its defining features (e.g., Hockett, 1963), sound symbolism allows one way for nonarbitrariness to play a role. It does this through congruencies between the sound symbolic associations of a word's phonemes and the word's meaning. An example of this could be when a word denoting something small contains phonemes that are sound symbolically associated with smallness (i.e., an instance of *indirect iconicity*, discussed later). These congruencies can have effects on language learning (e.g., Asano et al., 2015; Imai, Kita, Nagumo, & Okada, 2008; Perry, Perlman, & Lupyan, 2015; for a review, see Imai & Kita, 2014) and processing (e.g., Kanero, Imai, Okuda, Okada, & Matsuda, 2014; Lockwood & Tuomainen, 2015; Sućević, Savić, Popović, Styles, & Ković, 2015). Moreover, sound symbolic associations have also been shown to impact cognition more broadly, including effects on action (Parise & Pavani, 2011; Rabaglia, Maglio, Krehm, Seok, & Trope, 2016; Vainio, Schulman, Tiippana, & Vainio, 2013; Vainio, Tiainen, Tiippana, Rantala, & Vainio, 2016), memory (Lockwood, Hargoort, & Dingemans, 2016; Nygaard, Cook, & Namy, 2009; Preziosi & Coane, 2017), and categorization (Ković, Plunkett, & Westermann, 2010; Lupyan & Casasanto, 2015; for a recent review of sound symbolism effects, see Lockwood & Dingemans, 2015).

Interest in sound symbolism within psychology is on the rise. Ramachandran and Hubbard's (2001) article, which rekindled interest in the phenomenon,³ was one of only 28 published on sound symbolism and/or the closely related topic of iconicity (discussed later) in that year. For comparison, a total of 193 articles were published on sound symbolism and/or iconicity in 2016 (see Fig. 1). However, despite growing interest in the phenomenon, one topic that has largely been neglected is the mechanism underlying these associations. That is, mechanisms to explain *why* certain phonemes come to be associated with particular perceptual and/or semantic features. While there are a number of proposals, there is a scarcity of experimental work focused on adjudicating between them. One potential reason for this is that the mechanisms have yet to be thoroughly described

³ The topic itself dates back at least to the fifth century BC, when Plato's *Cratylus* takes place. This dialogue discusses the origin of words and contrasts a conventionalist perspective (i.e., that convention alone dictates the forms of words) with a naturalist perspective (i.e., that forms are naturally well suited for particular referents). These were popular topics of debate at the time (Sedley, 2013). It also includes interesting sound symbolic proposals, for instance *n/* being an internal sound, fit for meanings such as *within* or *inside*.

Table 1 Sample definitions of phonetic sound symbolism in the literature

Sound symbolism definitions

- “Sound symbolism is the process by which speakers link phonetic features with meanings non-arbitrarily” (D’Onofrio, 2013, p. 1).
 - “Synesthetic sound symbolism is the process whereby certain vowels, consonants, and suprasegmentals are chosen to consistently represent visual, tactile, or proprioceptive properties of objects, such as size or shape” (Hinton, Nichols, & Ohala, 1994, p. 4).
 - “Phonetic symbolism...proposes that an arbitrary linguistic sound itself carries symbolic weight, in that it evokes a sense of relatedness to other entities, such as color, touch, or emotion” (Hirata, Ukita, & Kita, 2011, p. 929).
 - “The idea of phonetic symbolism implies that sounds carry intrinsic symbolic connotations” (Koriat & Levy, 1977, p. 93).
 - “The term sound symbolism is used when a sound unit such as a phoneme, syllable, feature, or tone is said to go beyond its linguistic function as a contrastive, non-meaning-bearing unit, to directly express some kind of meaning” (Nuckolls, 1999, p. 228).
 - “Sound symbolism refers to cases in which particular images are associated with certain sounds” (Shinohara & Kawahara, 2010, p. 1).
-

and evaluated in a single work (though see Deroy & Auvray, 2013; Fischer-Jørgensen, 1978; French, 1977; Johansson & Zlatev, 2013; Masuda, 2007; Nuckolls, 1999; Shinohara & Kawahara, 2010); that is the aim of the present article. We begin by describing two well-known instances of sound symbolism to serve as reference points. Then, as an illustration of this topic's importance, the role of sound symbolism in language is reviewed. Next, we review the features of phonemes that may be involved in associations, and then explore the proposed mechanisms by which these features come to be associated with particular kinds of stimuli. Finally, we identify the outstanding issues that need to be addressed on this topic and suggest potential next steps for the field.

Size and shape symbolism

The two most well-known sound symbolic effects are typically traced to a pair of works from 1929 (though there are relevant earlier observations; e.g., Jespersen, 1922; von der Gabelentz, 1891; for a review, see Jakobson & Waugh, 1979). One of these is the *mil/mal effect* (Sapir, 1929), referring to an association between high and front vowels (see Table 2), and small objects; and low and back vowels, and large objects (Newman, 1933; Sapir, 1929). That is, when individuals are asked to pair nonwords such as *mil* and *mal* with a small and a large shape, most will pair *mil* with the small shape and *mal* with the large shape. Beyond a number of such explicit demonstrations (e.g., Thompson & Estes, 2011), the effect has also been shown to emerge implicitly.

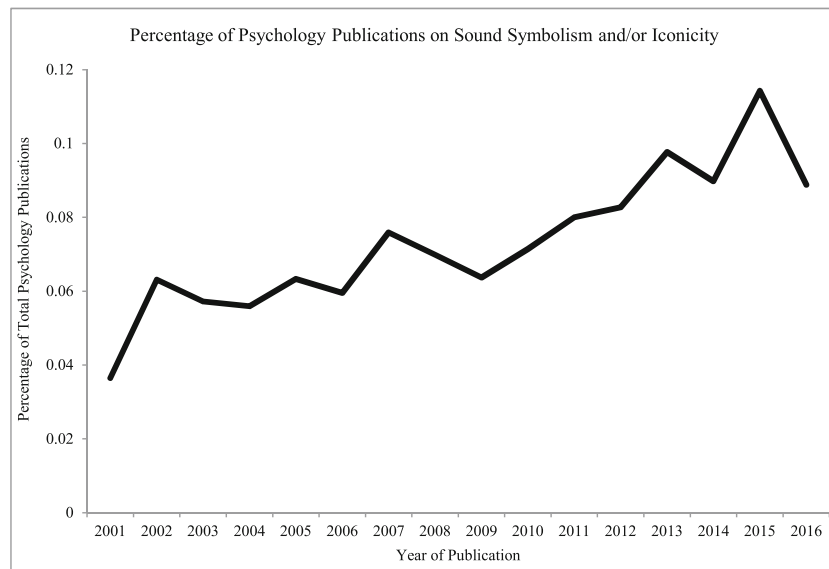


Fig. 1 Percentage of psychological publications per year that included the term *sound symbolism* and/or *iconicity*, according to PsycINFO

Participants are faster to respond on an implicit association task (IAT) if *mil*/small shapes and *mal*/large shapes share response buttons compared to when the pairing is reversed (Parise & Spence, 2012). In addition, participants are faster to classify a shape’s size if a sound-symbolically-congruent (vs. incongruent) vowel is simultaneously presented auditorily (Ohtake & Haryu, 2013). The effect has been demonstrated

across speakers of different languages (e.g., Shinohara & Kawahara, 2010) and at different points in the life span (e.g., in the looking times of 4-month-old infants; Peña, Mehler, & Nespor, 2011).

Another well-studied sound symbolic association is the *maluma/takete effect* (Köhler, 1929), referring to an association between certain phonemes and either round or sharp

Table 2 Definitions of linguistic terms used throughout the article (derived from Ladefoged & Johnson, 2010; Reetz & Jongman, 2009)

Phoneme term	Examples
Affricate consonants involve a combination of stops and fricatives.	/tʃ/ as in chat, /dʒ/ as in jack
Alveolar consonants involve the tip of the tongue contacting the alveolar ridge.	/t/ as in tab, /d/ as in dab
Approximant consonants involve a minor constriction in airflow that does not cause turbulence.	/l/ as in lack, /w/ as in whack
Back vowels are those articulated with the highest point of the tongue relatively close to the back of the mouth.	/u/ as in who’d, /ɑ/ as in hawed
Bilabial consonants involve the lips coming together in their articulation.	/m/ as in mat, /b/ as in bat
Fricative consonants involve a major constriction in airflow that does cause turbulence.	/f/ as in fat, /v/ as in vat
Front vowels are those articulated with the highest point of the tongue relatively close to the front of the mouth.	/i/ as in heed, /æ/ as in had
High vowels are those articulated with the tongue relatively close to the roof of the mouth.	/i/ as in heed, /u/ as in who’d
Low vowels are those articulated with the tongue relatively far from the roof of the mouth.	/æ/ as in had, /ɑ/ as in hawed
Nasal consonants involve airflow proceeding through the nose.	/m/ as in mat, /n/ as in gnat
Obstruent consonants involve a stoppage of, or turbulence in, the airflow; this includes stops, fricatives and affricates.	/p/ as in pat, /v/ as in vat, /tʃ/ as in chat
Rounded vowels are those articulated with rounded lips.	/u/ as in who’d, /oʊ/ as in hoed
Sonorant consonants involve no stoppage of, or turbulence in, the airflow; this includes nasals and approximants.	/m/ as in mac, /l/ as in lack
Stop consonants involve a stoppage of airflow.	/p/ as in pat, /b/ as in bat
Unrounded vowels are those articulated without rounded lips.	/i/ as in heed, /æ/ as in had
Velar consonants involve the back of the tongue contacting the soft palette.	/k/ as in cap, /g/ as in gap
Voiced consonants involve the vocal folds being brought close enough together to vibrate.	/b/ as in bam, /d/ as in dam
Voiceless consonants involve the vocal folds not being brought close enough together to vibrate	/p/ as in pat, /t/ as in tat

shapes. More recently, this has often been called the *bouba/kiki effect*, referring to the stimuli used by Ramachandran and Hubbard (2001) in their demonstration of the effect. In general, voiceless stop consonants (i.e., /p/, /t/, and /k/)⁴ and unrounded front vowels (e.g., /i/ as in *heed*) seem to be associated with sharp shapes; while sonorant consonants (e.g., /l/, /m/, and /n/), the voiced bilabial stop consonant /b/, and rounded back vowels (e.g., /u/ as in *who'd*), are associated with round shapes (D'Onofrio, 2013; Nielsen & Rendall, 2011; Ozturk, Krehm, & Vouloumanos, 2013; cf. Fort, Martin, & Peperkamp, 2014). As with the mil/mal effect, the maluma/takete effect has been repeatedly demonstrated using explicit matching tasks (e.g., Maurer et al. 2006; Nielsen & Rendall, 2011; Sidhu & Pexman, 2016). It also emerges on implicit tasks such as the IAT (Parise & Spence, 2012) and on lexical decision tasks, such that nonwords are responded to faster when presented inside of congruent (vs. incongruent) shape frames (e.g., a sharp nonword inside of a jagged vs. curvy frame; Westbury, 2005; cf. Sučević et al., 2015). It has been demonstrated in speakers of a number of different languages (e.g., Bremner et al., 2013; Davis, 1961; cf. Rogers & Ross, 1975) and in the looking times of 4-month-old infants (Ozturk et al., 2013; cf. Fort, Weiß, Martin, & Peperkamp, 2013; Pejovic & Molnar, 2016).

Arbitrariness and nonarbitrariness

Sound symbolism is relevant to our understanding of the fundamental nature of spoken language, in particular, to the relationship between the form of a word (i.e., its articulation, phonology, and/or orthography) and its meaning. One possibility is that this relationship is *arbitrary*, with no special connection between form and meaning (e.g., Hockett, 1963).⁵ Hockett (1963) described this lack of special connection as the absence of a “physical or geometrical resemblance between [form and meaning]” (p. 8). However this seems to only contrast arbitrariness with iconicity (see below). A more general way of characterizing this lack of a special connection is that aspects of a word's form cannot be used as cues to its meaning (Dingemanse, Blasi, Lupyan, Christiansen, &

Monaghan, 2015). As an illustration, it would be difficult to derive the meaning of the word *fun* from aspects of its form.⁶ An important related concept is *conventionality*, the notion that words only mean what they do because a group of language users have agreed upon a definition.

It is also possible for the relationship between form and meaning to be nonarbitrary, either through *systematicity* or *iconicity* (Dingemanse et al., 2015). Systematicity refers to broad statistical relationships among groups of words belonging to the same semantic or syntactic categories. For instance, Farmer, Christiansen, and Monaghan (2006) showed that English nouns tend to be more phonologically similar to other nouns than to verbs (and vice versa for verbs). Similarly, Reilly and Kean (2007) demonstrated that there are general differences in the forms of concrete and abstract English nouns. Importantly, systematicity does not involve relationships between words' forms and their specific meanings but broad relationships between groups of words and linguistic categories (Dingemanse et al., 2015). For instance, the nouns *member*, *prison*, and *student* are systematic in that they have a stress on their initial syllable (as do most disyllabic nouns; Sereno, 1986). This is a nonarbitrary property in that it is possible to derive grammatical category from word form. However, initial syllable stress is not related to these words' specific meanings in any particular way. While systematicity tends to occur on a large scale within a language, specific patterns of systematicity vary from language to language (Dingemanse et al., 2015).

The other way that language can be nonarbitrary is through iconicity: a resemblance between form and meaning.⁷ Instead of a holistic resemblance, this often emerges as a structural, resemblance-based mapping between aspects of a word's form and aspects of its meaning (Emmorey, 2014; Meir, Padden, Aronoff, & Sandler, 2013; Taub, 2001). For instance, consider the example used by Emmorey (2014), of the hand sign for *bird* in American Sign Language, in Fig. 2. Notice that specific features of the form map on to specific features of the meaning (e.g., the presence of a protrusion at the mouth, the ability of that protrusion to open vertically). Because only certain aspects of meaning are included in the mapping, there are elements of the concept bird that are not represented (e.g.,

⁴ Though note that the shape associations of the voiceless bilabial stop /p/ have been somewhat equivocal (see D'Onofrio, 2013; Fort et al., 2014).

⁵ Readers familiar with the sound symbolism and iconicity literature will no doubt notice the absence of reference to Ferdinand de Saussure's *Course in Linguistics* (1916), which famously stated that “the bond between the signifier and the signified is arbitrary” (p. 67). As reviewed in Hutton (1989), de Saussure may have intended to use the term *arbitrary* to describe the relationship between the abstract, mentalistic entities of the signifier and signified, as opposed to the form of a word and its referent in the world. It is this latter sort of arbitrariness that is relevant to sound symbolism. See Joseph (2015) for further discussion of this and de Saussure's later work, which explored iconicity as a factor in language change.

⁶ Surprisingly, finding a word to exemplify arbitrariness was quite difficult. This is illustrative of the point to follow, that most words contain a combination of arbitrary, systematic, and iconic elements. We chose *fun* because of its low iconicity rating (Winter, Perry, Perlman, & Lupyan, 2017) and derived systematicity value (Monaghan, Shillcock, Christiansen, & Kirby, 2014). Its length is also atypical of abstract nouns, which tend to be longer than concrete ones (Reilly & Kean, 2007), though this raises the interesting question of whether antisystematic words are arbitrary.

⁷ This discussion focuses on phonological iconicity; however, it is also possible to have iconicity at the level of morphemes (e.g., the addition of a plural suffix making a word larger; Jakobson, 1965), syntax (e.g., word order resembling temporal order; Perniss et al., 2010) and prosody (e.g., the tendency to use a faster rate of speech when discussing faster movements; Shintel, Nusbaum, & Okrent, 2006).



Fig. 2 The sign for *bird* in American Sign Language. Notice that specific aspects of the word's form map onto specific aspects of its meaning. For instance, the presence of a protrusion at the mouth, and the ability of that protrusion to open vertically. *Note.* From "ASLU," by W. Vicars, 2015 (<http://lifefprint.com/index.htm>). Copyright 1997 by Lifefprint Institute. Reprinted with permission.

its wings). An example of this in spoken language is onomatopoeia: words that sound like their referent. Take for instance the word *ding*, whose abrupt onset and fading offset map onto these features in the sound of a bell (Taub, 2001; see Assaneo, Nichols, & Trevisan, 2011). The preceding examples could be considered instances of *direct* iconicity, in which form maps directly onto meaning via resemblance (Masuda, 2007). This mapping is of course constrained by the form's modality; spoken language affords direct mapping onto meanings related to (or involving) sound, while signed languages are able to directly map onto spatial and kinesthetic meanings (Meir et al., 2013; Perniss, Thompson, & Vigliocco, 2010).

It is also possible for language to display *indirect* iconicity in which it is the *forms' associations* that map onto meaning (Masuda, 2007). This was put elegantly by Von Humboldt (1836) as cases in which sounds "produce for the ear an impression similar to that of the object upon the soul" (p. 73). In indirect iconicity it is the *impression* of the sound that maps onto meaning as opposed to the sound itself.⁸ Consider for instance the word *teeny* (/tini/). Because its meaning is not related to sound, its phonemes cannot map onto meaning directly. However, as mentioned, the high-front vowel

⁸ This kind of iconicity is by its very nature subjective, dependent on the associations a person makes (for a discussion see Hutton, 1989; Joseph, 2015). Nevertheless when an association is salient enough that it is apparent to a large group of language users, it merits consideration as a genuine phenomenon.

phoneme /i/ is sound symbolically associated with smallness. Thus, this phoneme maps onto smallness indirectly, by way of its sound symbolic association, allowing *teeny* to be indirectly iconic. This is the relevance of sound symbolism to language: it provides one mechanism by which words can be nonarbitrarily associated with their meanings.

The preceding examples of iconicity would be considered instances of *imagic* iconicity: a relationship between a single form and meaning (Peirce, 1974). However, some have proposed that sound symbolism plays a role in *diagrammatic* iconicity: cases in which the relationship between two forms resembles the relationship between their two meanings. Imagic and diagrammatic iconicity are sometimes referred to as *absolute* and *relative* iconicity, respectively (e.g., Dingemanse et al., 2015). Diagrammatic iconicity is often seen in ideophones, a class of words that depict various sensory meanings (beyond sounds) through iconicity (see Dingemanse, 2012). For instance, the Japanese ideophones *goro* and *koro* mean a heavy and a light object rolling, respectively. Note that *goro* begins with a voiced consonant while *koro* begins with a voiceless consonant; voiced (voiceless) consonants are associated with heaviness (lightness; Saji, Akita, Imai, Kantartzis, & Kita, 2013). Thus, the relationship between the sound symbolic properties of each word (i.e., one being sound symbolically heavier than the other) reflects the relationship between their meanings. At the moment it is unclear whether sound symbolism primarily contributes to indirect imagic iconicity or requires the comparison inherent in diagrammatic iconicity (e.g., Gamkrelidze, 1974). In Figure 6, in the Appendix, we propose a taxonomy of iconicity that is an attempt to synthesize the various distinctions that have been made in the literature.

There is a good deal of work demonstrating that iconicity is present in the lexicons of spoken languages.⁹ The clearest example of this is the widespread existence of ideophones. Although they are rare in Indo-European languages, they are common in many others, including sub-Saharan African languages, Australia Aboriginal languages, Japanese, Korean, Southeast Asian languages, South America indigenous languages, and Balto-Finnic languages (Perniss et al., 2010). Additionally, speaking to the psychological reality of ideophones, studies have shown that there are both behavioral (e.g., Imai et al., 2008; Lockwood, Dingemanse, & Hagoort, 2016) and neural differences (e.g., Kanero et al., 2014; Lockwood et al., 2016; Lockwood & Tuomainen, 2015) in the learning and processing of ideophones as compared to nonideophonic words (or ideophones paired with incorrect meanings).

There is also evidence that iconicity plays a role in the lexicon beyond ideophones. For instance, Ultan (1978) found

⁹ The presence of iconicity in signed languages is of course more obvious and less controversial (for a review, see Perniss et al., 2010).

that among languages that use vowel ablauting to denote diminutive concepts, most do so with high-front vowels. This is an example of indirect iconicity, occurring via high-front vowels' sound symbolic associations with smallness. In addition, Blasi, Wichmann, Hammarström, Stadler and Christiansen (2016) compared the forms of 100 basic terms across 4,298 languages and found, in addition to other patterns, that words for the concept *small* tended to include the high-front vowel /i/. Cross-linguistic studies have also reported evidence of indirect iconicity in, among other things, proximity terms (e.g., Johansson & Zlatev, 2013; Tanz, 1971), singular versus plural markers (Ullman, 1978), and animal names (Berlin, 1994). Additionally, the ability of individuals to guess the meanings of foreign antonyms at an above chance rate (e.g., Bankieris & Simner, 2015; Brown, Black, & Horowitz, 1955; Klank, Huang, & Johnson, 1971) has been attributed to indirect iconicity.

Taking an even broader view of language, Perry et al. (2015) and Winter et al. (2017) conducted large-scale norming studies in which 3,001 English words were rated on a scale with 0 indicating arbitrariness and 5 indicating iconicity. Many words had an average rating significantly greater than zero, indicating that this sample of words was not entirely arbitrary. Moreover, the iconicity of words in this sample is related to age of acquisition (Perry et al., 2015), frequency, sensory experience (Winter et al., 2017), and semantic neighborhood density (Sidhu & Pexman, 2017). Thus, instead of being a linguistic curiosity, iconicity appears to be a general property of language that behaves in a predictable manner, even in a less obviously iconic language such as English.

Of course, the existence of systematicity and iconicity does not discount the premise that arbitrariness is a fundamental property of language. As put by Nuckolls (1999), “throughout the exhaustive dissections and criticisms of the principle of arbitrariness, there has never been a serious suggestion that it be totally abandoned” (p. 246). Instead, arbitrariness, systematicity, and iconicity are seen as three coexisting aspects of language (Dingemanse et al., 2015). In fact, there is a growing appreciation that words do not fall wholly into the categories arbitrary and nonarbitrary but rather that individual words can contain both arbitrary and nonarbitrary elements (e.g., Dingemanse et al., 2015; Perniss et al., 2010; Waugh, 1992). For instance, consider the word *hiccups*. It is a noun with a stressed first syllable (a systematic property); it also imitates aspects of its meaning (an iconic property). However, without knowing its definition, one would not be able to fully grasp its meaning based solely on its form (an arbitrary property). It seems that each of these properties contribute to language in varying proportions; they each also provide unique benefits to language. That is, systematicity facilitates the learning of linguistic categories (e.g., Cassidy & Kelly, 1991; Fitneva, Christiansen, & Monaghan, 2009; Monaghan, Christiansen, & Fitneva, 2011). Iconicity makes

communication more direct and vivid (Lockwood & Dingemanse, 2015), and can facilitate language learning (e.g., Imai et al., 2008; for a review, see Imai & Kita, 2014). Lastly, decoupling form and meaning (i.e., arbitrariness) allows language to denote potentially limitless concepts (Lockwood & Dingemanse, 2015) and avoids confusion among similar meanings with similar forms (e.g., Gasser, 2004; Monaghan et al., 2011).

Phonetic features involved in sound symbolism

Before turning to a discussion of how sound symbolic associations between phonemes and particular stimuli arise, it is important to make clear that in the present review we conceptualize these associations as arising from associations between *specific phonetic features*¹⁰ and particular perceptual and/or semantic features. For instance, the association between high-front vowels and smallness (i.e., the mil/mal effect) is seen as arising from an association between some component acoustic or articulatory feature of high-front vowels, and smallness. Phonemes are multidimensional bundles of acoustic and articulatory features, any or all of which may afford an association with particular stimuli (e.g., Tsur, 2006). Indeed, Jakobson and Waugh (1979) opine that “most objections to the search for the inner significance of speech sounds arose because the latter were not dissected into their ultimate constituents” (p. 182). Thus, the first step is to delineate these various features of vowel and consonant phonemes that may be involved in associations.

Vowels are phonemes produced by changing the shape and size of the vocal tract through moving the tongue's position in the mouth and opening the jaw. This is done without obstructing the airflow—without having the articulators (e.g., tongue, lips) come together. The three main articulatory features that determine vowels' identity are: height (proximity of the tongue body to the roof of the mouth), frontness (proximity of the highest point of the tongue to the front of the mouth; see Fig. 3) and lip rounding.¹¹ Vowels are

¹⁰ We use the term *features* more broadly than it would be used in the context of a strict phonological analysis (e.g., Jakobson, Fant, & Halle, 1951). Our discussion of features is also less exhaustive than would be found in such a context.

¹¹ Another distinction is between tense (e.g., /i/ and /u/) and lax (e.g., /ɪ/ as in *hid*, and /ʊ/ as in *hood*) vowels. As noted by Ladefoged and Johnson (2010), this distinction is not simply based on muscular tension in their articulation; instead, the language-specific contexts in which they can appear differ. For instance, in English content words, tense vowels can appear in open syllables (e.g., *bee*, *boo*), while lax vowels cannot. While we have eschewed discussion of this in the main text in favor of dimensions that are more often discussed in the sound symbolism literature, there is some evidence of tenseness being involved in sound symbolism (e.g., Greenberg & Jenkins, 1966). Moreover, the tense/lax distinction is related to vowel length, with tense vowels tending to be longer than lax vowels (Ladefoged & Johnson, 2010); some studies have indeed implicated vowel length in sound symbolism (e.g., Newman, 1933).

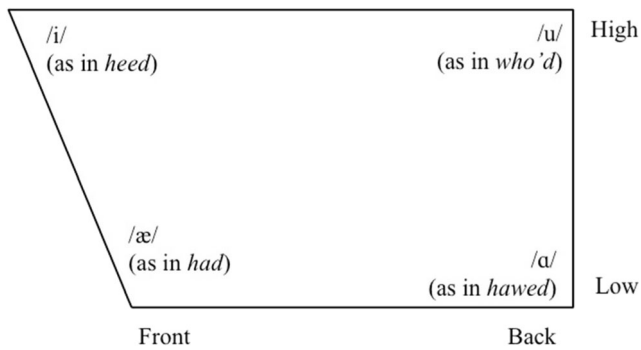


Fig. 3 An illustration of vowel space. The x -axis corresponds to the front–back dimension; the y -axis corresponds to the high–low dimension

described acoustically in terms of their formants: bands of high acoustic energy at particular frequencies. Tongue and jaw position serve to change the configuration of the vocal tract and affect which frequencies will resonate most strongly. The lowest of these formants (i.e., fundamental frequency) corresponds with the pitch of a vowel; it tends to be higher for high vowels, potentially because the tongue’s height “pulls on the larynx, and thus increases the tension of the vocal cords” (Ohala & Eukel, 1987, p. 207), increasing pitch. The next three formants determine the identity of a vowel. The frequency of the first formant (F1) is negatively correlated with height; the frequency of the second formant (F2) is positively correlated with frontness. These relationships are due to changes in the volume of resonating cavities in the vocal tract when the articulators are in different positions. In addition, lip rounding lowers the frequency of all formants above the first (in particular, the third). The distance between these formants (i.e., formant dispersion) is also important. For instance, front vowels are characterized by larger formant dispersion (between F1 and F2) than back vowels.

In articulating consonants, the airstream is obstructed in some way; consonants are defined based on the manner of this obstruction and the place where it occurs (Ladefoged & Johnson, 2010). Broadly speaking, consonants’ manner of articulation can be divided into obstruents (produced with a severe obstruction of airflow) and sonorants (produced without complete stoppage of, or turbulence in, the airflow; Reetz & Jongman, 2009). Obstruents include stops (in which airflow is entirely blocked and then released in a burst), fricatives (in which airflow is made turbulent by bringing two articulators together), and affricates (a combination of the two). Sonorants include nasals (in which airflow proceeds through the nasal cavity) and approximants (in which airflow is affected by bringing two articulators together, though not enough to create turbulence). In the case of obstruent consonants, they can also be distinguished by whether the vocal folds are brought close together enough to vibrate (i.e., voiced consonants) or not (i.e., voiceless consonants); sonorant consonants are typically voiced. Place of articulation refers to the location at which the airflow is affected, and especially

relevant categories include bilabials (in which the lips are brought together), alveolars (in which the tongue tip is brought to the alveolar ridge), and velars (in which the back of the tongue is brought to the soft palate).

As with vowels, each of these articulatory features of consonants have acoustic consequences. Stops involve a period of silence (potentially with voicing) followed by a burst of sound as they are released (potentially with aspiration). Fricatives cause turbulent noise in higher frequencies; nasals involve formants similar to vowels, though much fainter, while approximants have stronger formant structures.

Consonants and vowels also affect one another through coarticulation. That is, very few words involve a single phoneme. The gestures involved in producing sequences of phonemes are quick and result in adjacent sounds influencing the articulation of one another. For instance, vowels can affect consonants’ formant transitions (an acoustic cue to the place of articulation). In addition, a vowel’s pitch can be affected by the consonant that precedes it (e.g., higher when preceded by a voiceless obstruent; Kingston & Diehl, 1994).

Mechanisms for associations between phonetic and semantic features

Next we turn to the main topic of this review: how these phonetic features come to be associated with particular kinds of stimuli. This discussion will draw heavily from the literature on *crossmodal correspondences*, which, broadly speaking, can be defined as “the mapping that observers expect to exist between two or more features or dimensions from different sensory modalities (such as lightness and loudness), that induce congruency effects in performance and often, but not always, also a phenomenological experience of similarity between such features” (Parise & Spence, 2013, p. 792; also reviewed in Parise, 2016; Spence, 2011). For instance, individuals more readily associate bright objects with high-pitched sounds than with low-pitched sounds (Marks, 1974), and are faster to respond to objects if their brightness is congruent with a simultaneously presented tone (Marks, 1987). Our grouping of proposed explanations owes much to Spence’s (2011) grouping of proposed mechanisms for such crossmodal correspondences.

As noted by Parise (2016), the term *crossmodal correspondence* has been used to refer to associations between simple unidimensional stimuli, consisting of a single basic feature (e.g., pitch of pure tones, brightness of light patches) as well as associations between more perceptually complex, multidimensional stimuli, composed of multiple features from different modalities (e.g., linguistic stimuli, which contain multiple acoustic and articulatory features). If one considers crossmodal correspondences to encompass all associations between stimuli in different modalities, then sound symbolic

associations would certainly fall into this category (as in Parise & Spence, 2012; Spence, 2011). However, associations involving either simple or complex stimuli could potentially be distinct phenomena (see Parise, 2016). Thus, in the following review, we use the term *crossmodal correspondence* only to refer to associations between basic perceptual dimensions (e.g., brightness and pitch), which make up the majority of the term's usage (Parise, 2016). This draws a distinction between sound symbolic associations and crossmodal correspondences. Because phonemes are multidimensional stimuli, sound symbolism would be considered a distinct, though related, phenomenon from crossmodal correspondences. Thus, while mechanisms invoked to explain crossmodal correspondences can be informative, we must be cautious when extending them to sound symbolic associations.

In the following sections we group proposed explanations for sound symbolic associations into themes; note that although we think this grouping is helpful, there may be instances in which a given explanation could fit under multiple themes. Additionally, while we have included the themes that we feel best represent the existing literature, we acknowledge the possibility that other mechanisms may exist.

Mechanism 1: Statistical co-occurrence One mechanism proposed to explain associations between sensory dimensions is the reliability with which they co-occur in the environment (see Spence, 2011). That is, experiencing particular stimuli co-occurring in the world may lead to an internalization of these probabilities. This typically involves stimuli from a particular end of Dimension A tending to co-occur with stimuli from a particular end of Dimension B. One way of framing this is through the modification of Bayesian coupling priors, and the belief one has about the joint distribution of two sensory dimensions based on prior experience (Ernst, 2007).

Statistical co-occurrence has been proposed to explain the crossmodal correspondence between high (low) pitch and small (large) size (e.g., Gallace & Spence, 2006), due to the fact that smaller (larger) things tend to resonate at higher (lower) frequencies (see Spence, 2011). Another example is the association between high (low) auditory volume and large (small) size (e.g., Smith & Sera, 1992), which may arise from the fact that larger entities tend to emit louder sounds (see Spence, 2011). The plausibility of this mechanism has been demonstrated experimentally, by artificially creating co-occurrences between stimuli. Ernst (2007) presented participants with stimuli that systematically covaried in stiffness and brightness (e.g., for some participants, stiff objects were always bright). After several hours of exposure, participants demonstrated a crossmodal correspondence between these previously unrelated dimensions. Further evidence comes from a neuroimaging study that showed that after presenting participants with co-occurring audiovisual stimuli, the presentation of stimuli in one modality was associated with activity

in both auditory and visual regions (Zangenehpour & Zatorre, 2010).

This mechanism has been used to explain several sound symbolic associations. In these proposals, some component feature of the phonemes is claimed to co-occur with related stimuli in the environment. The most obvious application is to the mil/mal effect (see Spence, 2011). As mentioned, small (large) things tend to resonate at a high (low) frequency. Thus, front vowels may be associated with smaller objects because of front vowels' higher frequency F2. Similarly, the association between high vowels and smaller objects may be due to high vowels' higher pitch (Ohala & Eukel, 1987).¹² A similar explanation has also been applied to the association between front (back) vowels and short (long) distances (Johansson & Zlatev, 2013; Rabaglia et al., 2016; Tanz, 1971). Johansson and Zlatev (2013) noted that lower frequencies are able to travel longer distances and are therefore more likely to be heard from far away. Thus, we often experience more distant entities co-occurring with lower frequency sounds; this could potentially contribute to the association between back vowels (which have a lower F2) and long distance.

The mechanism of statistical co-occurrence has also been applied to *internally* experienced co-occurrences. For instance, Rummer, Schweppe, Schlegelmilch, and Grice (2014; also see Zajonc, Murphy, & Inglehart, 1989) proposed that some phonemes might develop associations with particular emotions due to an overlap between the muscles used for articulation and those used for emotional expression. Previous research suggested that simply adopting the facial posture of an emotion can facilitate experience of that emotion (i.e., the facial feedback hypothesis; Strack, Martin, & Stepper, 1988). Rummer et al. (2014) noted that articulating an /i/ involves contracting the zygomaticus major muscle which is also involved in smiling; conversely, articulating an /o/ (as in the German *hohe*) involves contracting the orbicularis oris muscle, which blocks smiling. They proposed that over time, the increased positive affect felt while articulating /i/ (due to facial feedback) will lead to that phoneme becoming associated with positive affect. Indeed, they showed that participants found cartoons funnier while articulating an /i/ as opposed to an /o/. However, they did not directly examine facial feedback as a mechanism. In addition, the validity of the facial feedback hypothesis has recently been called into question by failures to replicate Strack et al.'s original finding (Wagenmakers et al., 2016). Nevertheless, the notion that co-occurrences of

¹² These explanations contain an element of *indexicality*, one of Peirce's (1974) three sign elements, along with iconicity and symbolism (i.e., wholly arbitrary relationships). Indexes are defined by a relationship of contiguity between sign and object (e.g., smoke is an index of fire). Thus we might think of high frequencies being indexical of small size (see Johansson & Zlatev, 2013). This poses an interesting question regarding whether sound symbolism should only be discussed in relation to iconicity.

phonemes and *internal* sensations can lead to sound symbolic associations is a possibility that invites further evaluation.

One final statistical co-occurrence account is worth mentioning, despite the fact that it is not presented as an account of sound symbolism. Gordon and Heath (1998) reviewed findings that several vowel shifts (systematic changes in how vowels are articulated in a population) seem to be moderated by gender, with females leading raising and fronting changes and males leading lowering and backing changes. The term *raising*, for instance, refers to a given vowel being articulated with the tongue in a higher position than previously. They theorized that the different vocal tracts of women and men (contributing to women naturally having larger F2–F1 dispersion) might create an association between females and high-front vowel space (which has larger F2–F1 dispersion) and males and low-back vowel space. Females and males might then be drawn to gender stereotypical vowel space, leading to gender moderated vowel changes.¹³ Although the authors do not mention it, there is some evidence of a sound symbolic association between high-front vowels (low-back vowels) and femininity (masculinity; Greenberg & Jenkins, 1966; Tarte, 1982; Wu, Klink, & Guo, 2013; cf. Sidhu & Pexman, 2015). One might speculate that the natural co-occurrence between sex and formant dispersion contributes to this association.

There is a good deal of work that needs to be done to demonstrate that statistical co-occurrence is a viable mechanism for sound symbolism. The experimental evidence demonstrating that it can indeed create crossmodal correspondences (e.g., Ernst, 2007) makes it a promising mechanism. However, this evidence has been provided in the context of simple sensory dimensions; what remains to be seen is if such correspondences can then contribute to sound symbolic associations. That is, can a co-occurrence-based association between a component feature of a phoneme and certain stimuli create a sound symbolic association for that phoneme as a whole? One way to examine this question would be to present participants with isolated phonetic components (e.g., high vs. low frequencies) co-occurring with perceptual features (e.g., rough vs. smooth textures). Experimenters could then examine if this co-occurrence led to a sound symbolic association between phonemes containing said phonetic components (e.g., phonemes with a high vs. low frequency F2) and targets containing said perceptual feature (e.g., rough vs. smooth textures). Another approach would be to interfere with existing associations by presenting stimuli that contradict them (e.g., large objects making high-pitched noises) and then examining the effect on sound symbolic associations.

An important feature of this mechanism is that it requires experience, and thus assumes that at least some sound

symbolic associations are not innate (though, as will be discussed later, there are theories regarding evolved innate sensitivities to, and/or predispositions to acquire associations based on, certain statistical co-occurrences). As such, we might not expect associations that depend on statistical co-occurrences to be present from birth. Although Peña et al. (2011) found evidence for the mil/mal effect in four-month-old infants, it is possible that even these very young infants had already begun to gather statistical information about the environment (see Kirkham, Slemmer, & Johnson, 2002). Testing infants at an even younger age could allow us to investigate if less exposure to statistical co-occurrences results in a weaker sound symbolism effect (or the absence of an effect altogether). Of course, any differences between younger and older infants could simply be attributable to differences in cognitive development. Thus, another approach could be to test infants of the same age for associations based on co-occurrences that they are more or less likely to have experienced. For instance, young infants may have more experience of certain frequencies co-occurring with different sizes than with different distances; the effects of these differences in experience could be tested. Also, we would only expect associations of this kind to be universal if they are based on a universal co-occurrence. While natural co-occurrences reflecting physical laws (e.g., between pitch and size) may be relatively universal, it might be possible to find others that vary by location. For instance, some have speculated that advertising can create statistical co-occurrences that are relatively local, and that these potentially contribute to cultural variations in some crossmodal correspondences (e.g., Bremner et al., 2013). It could be informative to examine instances in which populations differ in culturally based statistical co-occurrences, and to compare their demonstrated associations. As mentioned by Wan et al. (2014), one might also consider effects of geographical differences (e.g., in landscape or vegetation) on statistical co-occurrences.

Mechanism 2: Shared properties Another broad class of accounts includes proposals that phonemes and associated stimuli may share certain properties, despite being in different modalities. Again, these properties in phonemes would likely derive from one or more of their component features. Individuals may then form associations based on these shared properties. These explanations can be divided into those involving low-level properties (i.e., perceptual) and those involving high-level properties (i.e., conceptual, affective, or linguistic).

Low-level properties Some perceptual features may be experienced in multiple modalities. For instance, one can experience size in both visual and tactile modalities. One way of explaining sound symbolic associations is to suggest that they involve an experience of the same perceptual feature in

¹³ In addition to this explanation, the authors do also mention the possibility of females and males being drawn to different areas of vowel space because of the sound symbolic associations of those areas.

both phonemes and associated stimuli. For instance, Sapir (1929; see also Jespersen, 1922) theorized that participants might have associated high vowels with small shapes in part because for high vowels the oral cavity is smaller during articulation. Thus, both phonemes and shapes had the property of *smallness*. Similarly, Johansson and Zlatev (2013) proposed this as one potential explanation for the association between high-front vowels and small distances. Many have also pointed out that the vowels associated with roundness (i.e., /u/, and /oʊ/ as in *hoed*) involve a rounded articulation (e.g., French, 1977; Ikegami & Zlatev, 2007; also suggested in Ramachandran & Hubbard, 2001). Note that these accounts involve some amount of abstraction or other mechanism by which features can be united across modalities, and do not necessarily imply that phonemes and stimuli possess *identical* perceptual features. Nevertheless, they do imply a certain amount of imitation between phonemes and associated features.

Others have proposed similar, though less direct, accounts. For instance, Saji et al. (2013) theorized that the association between voiced (voiceless) consonants and slow (fast) actions has to do with the shared property of duration. That is, in voiced consonants, the vocal cords vibrate prior to stop release, and thus for a longer time than in voiceless consonants. This longer duration might unite them with slow movements, which take a longer time to complete. Ramachandran and Hubbard (2001) also speculated that the maluma/takete effect might owe to an abruptness, or “sharp inflection” (p. 19) in both voiceless stops and sharp shapes. Indeed, voiceless stops involve a complete absence followed by an abrupt burst of sound; similarly, the outlines of sharp shapes involve abrupt changes in direction.

One final proposal is that a phoneme may be associated with body parts highlighted in its articulation (originally suggested by Greenberg, 1978). This account stands out from those discussed elsewhere in this review in that associations purported to derive from it have not been demonstrated experimentally but rather inferred from comparisons across languages. For instance, Urban (2011) found that across a sample of languages, words for *nose* and *lips* were more likely to contain nasals and labial stops, respectively, than a set of control words. In addition, Blasi et al. (2016) found that words for *tongue* tended to include the phoneme /l/ (for which the airstream proceeds around the sides of the tongue), while words for *nose* tended to include the nasal /n/. Importantly, the patterns documented by Blasi et al. did not seem to be a result of shared etymologies or areal dispersion; thus, the authors speculated that they could potentially have derived from sound symbolic associations (or a related phenomenon). If the association between phonemes and body parts that these findings seem to hint at exists, it would be much more direct and limited than other associations discussed in this review. Future behavioral studies might examine if, beyond these quasi-imitative relationships, phonemes are also associated with

stimuli that are *related* to the relevant body part¹⁴ (e.g., nasals and objects with salient odors). Such associations could ostensibly derive from the shared property of a salient body part.

High-level properties Others have proposed that the shared properties that produce sound symbolism are more *conceptual* in nature. For instance, L. Walker, Walker, and Francis (2012) suggested that crossmodal correspondences might emerge due to shared *connotative* meaning (i.e., what the stimuli suggest, imply, or evoke) among stimuli. Note that this is distinct from what the stimuli *denote* (i.e., what they directly represent). That is, a bright object denotes visual brightness, but this is distinct from a connotation of brightness, which can apply across modalities. For example, tastes and melodies can seem “bright.”

When we consider the fact that these suprasensory properties can be shared by stimuli across modalities, it becomes apparent that shared connotations might explain a wide variety of observed crossmodal correspondences. As an example, consider that high-pitched tones have the connotations of being brighter, sharper, and faster than low-pitched tones (L. Walker et al., 2012). These connotations of high-pitched tones might explain the association between high pitches and small stimuli (which also share these connotations). Moreover, P. Walker and Walker (2012; see also Karwoski, Odbert, & Osgood, 1942) proposed that there are a set of aligned connotations, such that a stimulus possessing one of them will also tend to possess the others. For instance, stimuli with the connotation of brightness will also tend to have connotations of sharpness, smallness, and quickness (L. Walker et al., 2012).

This framework may extend to sound symbolic associations (see P. Walker, 2016). That is, some sound symbolic associations might arise due to phonemes and stimuli sharing connotations. The connotations of phonemes would derive from the connotations of their component features. For instance, high-front vowels, which are high in frequency, have the same connotations as high frequency pure tones (e.g., brighter, sharper, faster). This might explain their association with small stimuli, which, as reviewed above, also share these connotations. In a test of this proposal, French (1977) hypothesized, and then investigated, a sound symbolic association between high-front vowels and coldness, based on a similarity in connotation between coldness and smallness. Indeed, his participants reported that nonwords containing the vowel /i/ were the “coldest” while those containing /a/ (as in *hawed*) were the “warmest.”

Similar explanations have also been applied to shape sound symbolism. Bozzi and Flores D’Arcais (1967) asked

¹⁴ The phonestheme *sn-*, which appears in words related to the nose and mouth (e.g., *snarl*, *sneeze*, *sniff*; see the Language Patterns section), may be indicative of such an association (e.g., Waugh, 1993).

participants to rate compatibility between nonwords and shapes, and also to rate both kinds of stimuli on semantic differential scales (i.e., Likert scales anchored by polar adjectives, used to measure connotations). They found that compatible nonwords and shapes tended to have similar connotations (e.g., sharp nonwords and shapes were both rated as being fast, tense, and rough). Gallace, Boschin, and Spence (2011) made a similar proposal to explain their finding that round and sharp nonwords were differentially associated with certain tastes. They found that these associations were predicted by similar ratings of nonwords and tastes on connotative dimensions such as tenseness or activity.

A limitation of this account is that it begs the question of how phoneme features come to be associated with their connotations. There are also several conceptual clarifications required. For instance, in cases of several shared connotations, is one primary in creating the association? In addition, there is a need to clarify the distinction between a given phoneme's connotations and its sound symbolic associations. That is, when participants rate a given vowel as belonging to the "small" end of a large/small semantic differential scale, does that describe a connotation, an associated perceptual (i.e., denotative) feature, or both? Should connotations themselves be considered instances of sound symbolism? The exact connotative dimensions involved also require further elaboration. Much of Walker's work focuses a core set of connotations including: light/heavy, sharp/blunt, quick/slow, bright/dark, and small/large (e.g., P. Walker & Walker, 2012). Others have focused on connotations that comprise the three factors of connotative meaning discovered by Osgood, Suci, and Tannenbaum (1957),

namely, evaluation (e.g., good/bad), potency (e.g., strong/weak), and activity (e.g., active/passive; e.g., Miron, 1961; Tarte, 1982).

It has also been proposed that some crossmodal correspondences arise via transitivity (e.g., Deroy, Crisinel, & Spence, 2013). That is, if there exists an association between Dimensions A-B, and B-C, this might create an association between Dimensions A-C. French (1977) made a similar suggestion for sound symbolic associations. He theorized that phonemes are only *directly* related with a small number of stimulus dimensions, and that these mediate relationships with other stimulus dimensions. For instance, high-front vowels may be directly associated with smallness, which mediates a relationship between high-front vowels and other properties related to smallness (e.g., thinness, lightness, quickness). A clarification to make going forward is whether these mediated effects involve relationships among denotative (as in Deroy et al., 2013) or connotative dimensions, or both (see Fig. 4).

A related proposal is that stimuli may be associated by virtue of having the same *impression* on a person. That is, instead of being united through a shared conceptual property, stimuli may be associated because they have a similar effect on a person's level of arousal or affect (Spence, 2011). Indeed there is some evidence of hedonic value (Velasco, Woods, Deroy, & Spence, 2015) and associated mood (Cowles, 1935) underlying crossmodal correspondences. This account has not yet been examined in the context of sound symbolism. However, as is discussed elsewhere in this review, there has been some work proposing a link between phonemes and particular affective states (e.g., Nielsen & Rendall, 2011, 2013; Rummer et al., 2014).

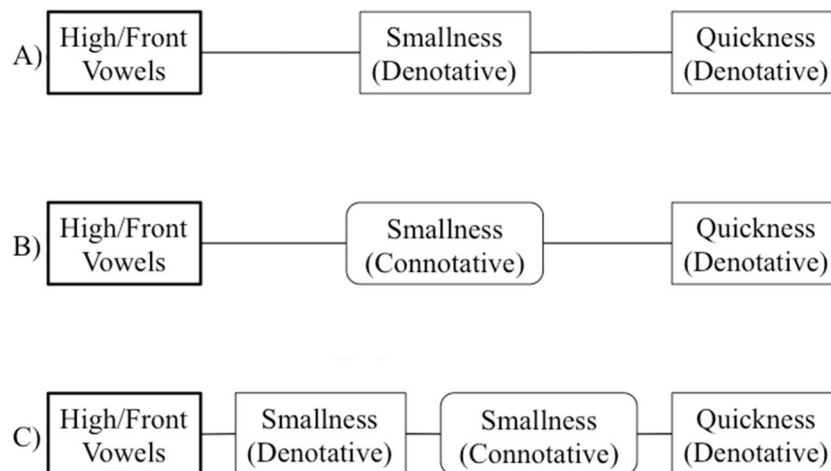


Fig. 4 Different ways in which denotative and/or connotative dimensions may result in mediated relationships. **a** The sort of mediation discussed by Deroy et al. (2013), in which the perceptual dimension of smallness might mediate a relationship between high/front vowels and the perceptual dimension of quickness via transitivity. **b** A relationship based off of P. Walker and Walker (2012; which might be considered a mediated one) in which the connotation of smallness might

mediate a relationship between high/front vowels and the perceptual dimension of quickness. **c** An example of mediation involving both denotative and connotative dimensions, in which high/front vowels are associated with the perceptual dimension of quickness because of the phonemes' association with the perceptual dimension of smallness, and that dimension's association with the perceptual dimension of quickness (via a shared connotation)

Lastly, some have theorized that crossmodal correspondences arise when the two dimensions share the same labels (e.g., Martino & Marks, 1999). For instance, the correspondence between pitch and elevation may derive from the use of the labels *high* and *low* for both. Evidence for this has come from the fact that speakers of languages using different labels for pitch (e.g., *high/low* in Dutch; *thin/thick* in Farsi) show different crossmodal correspondences (e.g., height and pitch in Dutch speakers; height and thickness in Farsi speakers; Dolscheid, Shayan, Majid, & Casasanto, 2013). Although this has not yet been proposed for sound symbolic associations, there are some relevant observations. For instance, front and back vowels are sometimes referred to as *bright* and *dark* vowels, respectively (e.g., Anderson, 1985). This corresponds to the visual stimuli with which either group of phonemes is associated (Newman, 1933). However, this example is only intended to serve as an illustration; at the moment, the relevance of this account to sound symbolism is purely speculative. In addition, a question related to this general explanation is one of directionality: do shared linguistic labels create associations, or vice versa, or both? Dolscheid et al. (2013) demonstrated that teaching Dutch speakers to refer to pitch in terms of thickness led to effects that resembled those of Farsi speakers. Speaking to the converse, Marks (2013) discussed the notion that crossmodal correspondences might contribute to the creation of linguistic metaphors, and the use of a term from one sensory modality to describe sensations in another (see also Shayan, Ozturk, & Sicoli, 2011).

An important step in testing theories based on shared properties will be demonstrating the involvement of the hypothesized shared properties (see Table 3 for a summary of properties). With regard to conceptual properties, a potential starting point could be to examine ratings on semantic differential scales for phonemes and associated stimuli, to test whether they indeed share connotations. The next step could be to verify activation of the shared conceptual properties, potentially by examining if they become more accessible following sound symbolic matching. For affect-based associations, one could examine whether phonemes and associated stimuli elicit comparable changes in a person's self-reported mood. Deroy et al. (2013) theorized that mediated relationships (involving denotative dimensions) should be weaker than direct ones. This provides a potential way of detecting such relationships. In addition, several of these theories may depend on developmental milestones (e.g., the acquisition of language) and thus make different predictions for sound symbolic effects when individuals are tested before and after these milestones are reached. Lastly, associations based on shared properties would be expected to vary cross-culturally to the extent that stimuli differ in their associated properties across cultures.

An outstanding question that is important for these accounts is whether participants only recognize shared properties and form associations when asked to do so during a task.

For instance, when asked to rate the similarity between non-words and tastes, participants might very well consider properties that the two have in common. However, this does not mean that such associations exist outside of that task context. One could address this issue by examining whether associations are detectable using implicit measures (e.g., priming) that do not force participants to consider the relationships between stimuli in an overt way. Indeed, P. Walker and Walker (2012) demonstrated that a crossmodal correspondence based on connotation could affect responses on an implicit task.

Mechanism 3: Neural factors The third mechanism includes proposals that sound symbolic associations arise because of structural properties of the brain, or the ways in which information is processed in the brain. To be clear, this is not to imply that other mechanisms do not rely on neural factors. The difference here is that the following theories propose neural factors to be the proximal causes of the associations.

A theory described in the crossmodal correspondence literature suggests that there may be a common neural coding mechanism for stimulus magnitude, regardless of modality. For instance, Stevens (1957) noted that increases in *stimulus intensity* result in higher neuronal firing rates. In a similar vein, Walsh (2003) proposed that a system in the inferior parietal cortex is responsible for coding *magnitude*, again across modalities. Thus, for stimulus dimensions that can be quantified in terms of more or less (e.g., more or less loud, more or less bright), this common neural coding mechanism may lead to an association between the “more” and the “less” ends of each dimension (see Spence, 2011). For instance, the correspondence between high (low) volume and bright (dim) objects (Marks, 1987) may have to do with the fact that they are both high (low) in magnitude (see Spence, 2011). So far this has not been extended to sound symbolic associations. However, it may be a viable mechanism when involving phonetic features that can be characterized in terms of magnitude.¹⁵

Another relevant theory is based on a hypothesized relationship between the brain regions associated with grasping and with articulation. Some have proposed that the articulatory system originated from a neural system responsible for grasping food with the hands and opening the mouth to receive it, resulting in a link between articulation and grasping (see Gentilucci & Corballis, 2006). Vainio et al. (2013) demonstrated that participants were faster to make a precision grip (i.e., thumb and forefinger) while articulating the phonemes /t/ or /i/, and faster to make a power grip (i.e., whole hand) while

¹⁵ Of course, it is possible that this magnitude matching is not neurally based. For instance, Marks (1989) theorized that loud and bright stimuli might share a semantic code (i.e., be represented as *intense*). Thus, magnitude matching might be conceptualized as being based on the shared conceptual properties of high intensity or low intensity, as opposed to being fundamentally neural in origin.

Table 3 A summary of the shared properties that could be involved in sound symbolism

Shared Property	Example
Perceptual feature	High-front vowels and small shapes sharing the property <i>smallness</i> (Sapir, 1929)
Magnitude or intensity ^a	Both high volume and brightness being high in magnitude (Spence, 2011)
Connotation	Stop consonants and angular shapes having the connotations of being fast and tense (Bozzi & Flores D'Arcais, 1967)
Relationship with a mediating dimension	High-front vowels being associated with thinness via the mediating dimension of size (French, 1977)
Affective quality and resulting impression	Sweet taste and round shape being united via their positive hedonic value (Velasco et al., 2015)
Linguistic label	Vowels referred to as <i>bright</i> or <i>dark</i> being associated with high and low brightness, respectively.

^a Magnitude and intensity are discussed in the section on neural factors

articulating the phonemes /k/ or /a/. Note that the articulation of each set of phonemes reflects the performance of either kind of grip.¹⁶ Vainio et al. theorized that the mil/mal effect might emerge from these associations (see also Gentilucci & Campione, 2011). For instance, seeing a small shape may elicit the simulation of a precision grip (Tucker & Ellis, 2001), which would then also activate a representation of the phoneme /i/'s articulation. It should be noted, however, that a follow up study by this group found participants were no faster to articulate an /i/ (/a/) in response to a small vs. large (large vs. small) target (Vainio et al., 2016).¹⁷ Thus, there is still a need for more direct evidence of the proposed links.

An ideal way to examine these neural theories would be to use neuroimaging. For instance, it would be informative to test for activation in the hypothesized magnitude-coding region when processing phonemes and related stimuli. Likewise, testing for activation in motor regions associated with articulation, in response to graspable objects, could also provide insight into articulation/grasping as a neural mechanism. There is recent evidence for the converse relationship: increased activity in motor regions associated with performing a precision or power grip, while articulating /ti/ or /ka/, respectively (Komeilipoor, Tiainen, Tiippana, Vainio, & Vainio, 2016). These mechanisms should be largely universal, and thus the neural accounts predict that sound symbolic associations should not be modulated by culture.

Mechanism 4: Species-general associations Some have explained sound symbolism as based on species-general, inherited associations. While other mechanisms may involve

¹⁶ The authors theorize that two separate, but potentially related, processes may be at work. The links between vowels and grips may be due to double grasp neurons: the mouth prepares to receive an object whose size is related to hand grip size. The links between consonants and grips may be due to a tendency to mirror hand movements with the speech musculature; for instance, note the similarity between the articulation of /t/ and a precision grip.

¹⁷ Interestingly though, participants were faster to articulate an /m/ or a /n/, in response to a round or a sharp shape, respectively.

evolved processes, the following theories propose that the associations themselves (as opposed to the processes leading to those associations) are a result of evolution.

One of the most widely cited explanations for the mil/mal effect is Ohala's (1994) frequency code theory. This is based on the observation that many nonhuman species use low-pitched vocalizations when attempting to appear threatening, and high-pitched vocalizations when attempting to appear submissive or nonthreatening (Morton, 1977). Ohala proposes that these vocalizations appeal to, and are indicative of, an innate cross-species association between high (low) pitches and small (large) vocalizers (viz. the frequency code). Thus, when an animal wants to appear threatening, they use a low-pitched vocalization in order to give off an impression of largeness. Ohala theorizes that humans' association between frequency (e.g., in vowels' fundamental frequency and F2) and size is due to this same frequency code. At a fundamental level, this explanation is based on co-occurrence (i.e., between pitch and size); however, it is argued that sensitivity to this co-occurrence has become innate. As evidence for this innateness, Ohala points to the fact that male voices lower at puberty: precisely when they will need to use aggressive displays (i.e., low-pitched vocalizations) to compete for a mate. He argues that such an elaborate anatomical evolution would only have been worthwhile if it appealed to an innate predisposition in listeners. Nevertheless, Ohala concedes that the frequency code may require some postnatal experience of relevant environmental stimuli, to be fully developed. Thus, one might regard the frequency code hypothesis as an innate predisposition to develop an association, rather than as an innate association per se.

It is important to note that while many studies have found a relationship between fundamental frequency and body size in several species (e.g., Bowling et al., 2017; Charlton & Reby, 2016; Gingras, Boeckle, Herbst, & Fitch, 2013; Hauser, 1993; Wallschläger, 1980), others have not (e.g., Patel, Mulder, & Cardoso, 2010; Rendall, Kollias, Ney, & Lloyd, 2005; Sullivan, 1984). As noted by Bowling et al. (2017), a relevant factor seems to be the range in body sizes

studied, with more equivocal effects when studying the relationships within a given category than across categories (e.g., within a species vs. across species; cf. Davies & Halliday, 1978; Evans, Neave, & Wakelin, 2006). In response to these equivocal findings, Fitch (1997) presented results from research with rhesus macaques, demonstrating that formant dispersion may be a better indicator of body size than fundamental frequency. It is beyond the scope of this review to adjudicate between these two cues. However, to the extent that formant dispersion is a more reliable cue to size than fundamental frequency, the frequency code hypothesis may require reframing. It is relevant to note that the mil/mal effect can be characterized in terms of formant dispersion, which is larger for front vowels than back vowels, and decreases from high-front vowels to low-front vowels.¹⁸

In a similar vein, Nielsen and Rendall (2011, 2013) note that many nonhuman species use harsh punctuated sounds in situations of hostility and high arousal; and smoother, more harmonic sounds in situations of positive affiliation and low arousal.¹⁹ Notably, the meanings of these calls do not need to be learned by conspecifics, suggesting an innate sensitivity to their meanings (Owren & Rendall, 2001). There is also evidence of this in humans: infants use harsh (smooth) sounds in situations of distress (contentment); adults use harsh and punctuated voicing patterns in periods of high stress (Rendall, 2003). Nielsen and Rendall theorize that the evolved semantic-affective associates of these two types of sounds may extend to phonemes with similar acoustic properties: namely obstruents and sonorants. For instance, swear words (which can be considered threatening stimuli) contain a relatively large proportion of obstruents (Van Lancker & Cummings, 1999). This could contribute to the maluma/takete effect, and to associations between stop phonemes (sonorant phonemes) and sharp (round) shapes. Such an account would depend on sharp shapes seeming more dangerous than round shapes, and indeed there is some speculation in this regard (Bar & Neta, 2006).

A potential limitation of the claims regarding evolved and/or innate traits is the challenge of generating testable hypotheses from these accounts. One approach would be to examine whether the relevant associations are present universally, and from a very young age. While there is evidence for sensitivity to the mil/mal effect (Peña et al., 2011) and the maluma/takete effect (Ozturk et al., 2013) in four month-old infants, it is

notable that two other studies have failed to find evidence of infant sensitivity to the maluma/takete effect at that age (Fort et al., 2014; Pejovic & Molnar, 2016). In addition, one might debate whether observing an effect at four months of age is sufficient to infer its innateness. Thus, the evidence for innateness is not overwhelming at present. At least one crossmodal correspondence has been demonstrated in infants between 20 and 30 days old (Lewkowicz & Turkewitz, 1980), and it would be informative for future studies to examine sensitivity to sound symbolism at a similar age. Another approach could be a comparative one, examining if non-humans demonstrate sound symbolism. Ludwig, Adachi, and Matsuzawa (2011) reported a crossmodal correspondence between pitch and brightness in chimpanzees, suggesting that such an investigation might be worthwhile.

Mechanism 5: Language patterns One final group of theories proposes that sound symbolic associations emerge due to patterns in language. This is, of course, related to the first mechanism discussed (i.e., statistical co-occurrence); the important distinction is that, as opposed to observing co-occurrences in the environment, the theories to be discussed propose sound symbolic associations might derive from co-occurrences between phonological and semantic features in language. An example of this would be associations derived from phonesthemes: phoneme clusters that tend to occur in words with similar meanings (e.g., *gl-* in words relating to light, such as *glint*, *glisten*, *glow*; see Bergen, 2004). After repeated exposure, individuals might come to associate /gl/ with brightness, for instance. Indeed there is evidence of individuals using their knowledge of phonesthemes when asked to generate novel words (e.g., using the onset *gl-* when asked to create a nonword related to brightness; Magnus, 2000). Bolinger (1950) even suggested that phonesthemes may “attract” the meanings of semantically unrelated words that contain the relevant phoneme clusters, leading to semantic shifts towards the phonesthemic meaning.

Such proposals are typically presented as an explanation for a distinct subset of sound symbolism, and not as an explanation for sound symbolism as a whole (e.g., Hinton et al., 1994). Indeed, our operational definition would consider associations arising in this manner to be a separate phenomenon altogether. Nevertheless, some have proposed that language patterns can explain *all* of sound symbolism (e.g., Taylor, 1963). This kind of proposal has, however, not been supported by large-scale corpus analyses. For instance, a study by Monaghan, Mattock, and Walker (2012) did not find overwhelming evidence that certain phonemes tend to occur in meanings related to roundness or sharpness. This would seem to suggest that the maluma/takete effect cannot be explained by language patterns. We described some other instances of indirect iconicity in the lexicon earlier in this paper, but the fact that many of these instances emerge across large samples

¹⁸ Ohala (1994) mentions that vowel sound symbolism may depend on formant dispersion, citing Fischer-Jørgensen (1978), potentially suggesting that Ohala saw the frequency code theory as compatible with a focus on formant dispersion.

¹⁹ These two accounts do seem to be related. As noted by Morton (1994), “aggressive animals utter low-pitched often harsh sounds...appeasing animals use high-pitched, often tonal sounds” (p. 350).

of languages leads to the conclusion that they are the result of sound symbolism as opposed to the cause of it (e.g., Blasi et al., 2016; Wichmann, Holman, & Brown, 2010).

There is, however, support for a weaker version of this claim, namely, that language patterns modify and constrain sound symbolism. For instance, Imai and Kita (2014) proposed that young infants are sensitive to a wide variety of sound symbolic associations, but that associations not supported by the phonology of an infant's language, or inventory of sound symbolic words, tend to fade away as the infant develops. This proposal is supported by evidence of a greater sensitivity to *foreign* sound symbolic words in children as compared to adults (Kantartzis, 2011). There is also evidence of a language's phonology moderating sound symbolic associations for speakers of that language. A basic example of this is the finding that individuals perceptually assimilate phonemes that do not appear in their language into ones that do (e.g., Tyler, Best, Faber, & Levitt, 2014; see Best, 1995). Sapir (1929) theorized that this may have been the reason English speaking participants did not rate nonwords containing /e/ (as in the French *été*) as being as small as expected. Because this phoneme does not appear in English, participants may have projected onto it the qualities of the diphthong /e/ (as in *hay*), which begins lower than /e/ for many speakers (Ladefoged & Johnson, 2010). Another example comes from a study by Saji et al. (2013), who found that high-back vowels were associated with slowness by Japanese speakers but with quickness by English speakers. They theorized that this was because this vowel is rounded in English but not in Japanese. Lastly, there is recent evidence that the distributional properties of phonemes in a given language can impact their tendency to show sound symbolic associations for speakers of that language (i.e., less frequent phonemes may be more likely to have sound symbolic associations; Westbury, Hollis, Sidhu, & Pexman, 2017).

Contextual factors One final topic that deserves mention is the role of various contextual factors in sound symbolism. As in the weaker version of the language patterns theory outlined above, contextual factors likely moderate the *expression* of sound symbolic associations rather than create them. For instance, some have theorized that forced-choice tasks may lead participants to become aware of shared properties among stimuli that they would not have considered otherwise (e.g., Bentley & Varron, 1933; French, 1977). In addition, some authors have speculated that pairing sounds with congruent meanings in real language may serve to highlight potential associations (e.g., Waugh, 1993). Dingemanse et al. (2016) point out that in some cases it is necessary to know the definition of a word in order to appreciate the sound symbolic association between its phonemes and meaning. That is, would one

appreciate the sound symbolism of *goro* without knowing that its definition related to heaviness? Tsur (2006) characterizes sound symbolic associations as “meaning potentials” (p. 917) that can be actualized by associating phonemes with meanings in language. As noted by Werner and Kaplan (1963), sounds demonstrate plurisignificance, in that they are able to be associated with multiple different dimensions. Tsur suggests that the semantic context in which words appear might highlight some potential associations over others. Lastly, prosody has been theorized to direct individuals towards particular sound symbolic associations (Dingemanse et al., 2016).

Another potential factor to consider is cultural variation in conceptualizations of the relationship between sound and meaning. Nuckolls (1999) reviewed case studies of a number of societies in which language sounds are seen as intimately related to the external world. For instance, the Navajo view air as a source of life, and manipulating that air in the service of creating linguistic sound as one way of making “contact with the ultimate source of life” (Reichard, 1944, 1950; Witherspoon, 1977, p. 61). As another example, different states of water (e.g., swirling, splashing) represent important landmarks for the Kaluli people of Papua New Guinea (Feld, 1982). Their language contains a number of ideophones that depict these different states of water, representing a fascinating interplay of linguistic sound and geography. This interplay is exemplified in their poetry, which depicts waterways in both sound and structure. Indeed, some have speculated that variations in ideophone usage may result from cultural variation in cognitive styles (e.g., Werner & Kaplan, 1963). One wonders if cultural factors may moderate the expression of sound symbolic associations.

Outstanding issues and future directions

We have outlined five mechanisms that have been proposed to explain sound symbolic associations: the features of the phonemes co-occurring with stimuli in the environment, shared properties among phoneme features and stimuli, overlapping neural processes, associations created by evolution, and patterns extracted from language. There are a number of outstanding issues in the literature, and it is to these that we now turn our attention.

Phonetic features So far in this review we have been equivocal on whether sound symbolism involves acoustic or articulatory features. In fact, there is no need to attribute the phenomenon to one or the other; most theorists allow for both to potentially play a role (e.g., Newman, 1933; Nuckolls, 1999; Ramachandran & Hubbard, 2001; Sapir, 1929; Shinohara & Kawahara, 2010; Westermann, 1927). This is commensurate

with the notion of phonemes as bundles of acoustic and articulatory features, either/both of which can be associated with targets in sound symbolism (e.g., Tsur, 2006).²⁰ Indeed there is evidence of both playing a role. For instance, Tarte's (1982) research comparing vowels to pure tones showed that vowels are associated with some stimuli in a way that would be expected if pairings were based on vowels' component frequencies. Eberhardt's (1940) discovery of sound symbolism in profoundly deaf individuals suggested that articulatory features in isolation can contribute to sound symbolism (though admittedly in a specific population; cf. Johnson, Suzuki & Olds, 1964).

While it seems reasonable to assume that both articulatory and acoustic features can play a role in sound symbolism, a potential topic for future research could be examining their relative contributions to particular effects. It may be that some associations are more dependent on articulatory features while others are more dependent on acoustic features. Of course, because acoustic and articulatory features are often inextricably linked (i.e., changes in articulation often result in changes in acoustics), this is an extremely difficult question to address. Even presenting linguistic stimuli visually for silent reading, or auditorily for passive listening, would not be sufficient to isolate acoustic features, as studies have shown that these can both lead to covert articulations (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003). Moreover, as pointed out by Eberhardt (1940), even acoustic features such as frequency can have tactile properties (i.e., felt vibrations). Nevertheless, because mechanisms of association are often based on particular features (e.g., the statistical co-occurrence of *acoustic frequency* and *size*), pinpointing the features involved could help adjudicate between potential mechanisms' roles in a certain effect. Future research might examine this by manipulating the component frequencies of vowels while maintaining their identity, or interfering with covert articulations, and then observing the effect on specific associations. In addition, beyond simply comparing the relative weighting of acoustic and articulatory features as a whole, it will be important to also consider the relative weighting among various acoustic features and articulatory features.

A related question is how individuals navigate the various associations afforded by phonemes' bundle of features. For instance, what leads individuals to weigh certain phoneme features more heavily than others? Recall that the phoneme /u/ is associated with largeness (Newman, 1933). This seems to suggest that individuals place more emphasis

on the association afforded by this phoneme's features as a back vowel (i.e., largeness) than as a high vowel (i.e., smallness).²¹ The matter is further complicated by the possibility that a given feature can afford associations with different ends of the same dimension. As an illustration, consider Diffolth's (1994) observation that the Bahnar language contains associations between high vowels and largeness (which contrasts with the typical mil/mal effect). Diffolth theorized that this resulted from a focus on the amount of space that the tongue takes up in the vocal tract (larger for high vowels), as opposed to the amount of space left empty (smaller for high vowels). Thus, this articulatory feature might potentially afford different (and conflicting) associations. Of course, this begs the question of why certain potential associations are more commonly observed than others (Nuckolls, 1999). Understanding what leads to the formation of certain associations out of the myriad of possibilities is an important topic for future research. This not only includes associations on an individual level, but also the crystallization of these associations in a given lexicon (in cases of indirect iconicity).

Lastly, it is worth briefly considering the role of *visual* features in sound symbolic effects. One example is the letters used to code for phonemes; for instance, visual features are sometimes presented as an important contributor to the maluma/takete effect (see Cuskley, Simner, & Kirby, 2015). In fact, Cuskley et al. (2015) showed that the visual roundness/sharpness of letters was a stronger predictor of nonword-shape pairing than was consonant voicing. However, given that sound symbolic effects emerge in a culture without a writing system (Bremner et al., 2013), in pre-literate infants (Ozturk et al., 2013; Peña et al., 2011; cf. Fort et al., 2013; Pejovic & Molnar, 2016), with learned neutral orthographies (Hung et al., 2017), and are not affected by direct manipulations of font (Sidhu, Pexman, & Saint-Aubin, 2016), it seems probable that orthography is, as the very least, not the sole contributor to these effects. Nevertheless the contribution of orthography relative to those of acoustics and articulation is still an open question. If orthographic features were found to play a large role in sound symbolism, it might weaken the claims of some theories that rest on phonological and/or articulatory features (e.g., the frequency code hypothesis, double grasp neurons). Associations based in orthography would likely be due to shared low-level perceptual features among letters and associated stimuli (though for potential roles of connotation, see Koriatic & Levy, 1977; P. Walker, 2016). In addition, some articulatory features have very strong visual cues (e.g., lip rounding). It remains to be seen if it is possible to separate these features from the tactile properties of articulation.

²⁰ With this in mind, *sound symbolism* becomes something of a misnomer, as it seems to imply that acoustic features drive associations. *Phonetic symbolism*, a term that is sometimes used to refer to the same effect (see Spence, 2011; Table 1) might be more appropriate. However, we elected to use *sound symbolism* since it is the more common term (e.g., used exclusively 45 times in 2016, compared to 12 for *phonetic symbolism*, per PsycINFO).

²¹ Though this might be part of the reason why Newman (1933) found that /u/ was not rated as large as /ɔ/ (a mid-back vowel), for instance.

Relationship with crossmodal correspondences An open question is the extent to which discoveries regarding crossmodal correspondences can be applied to sound symbolism. In particular, one might wonder how well mechanisms of association for crossmodal correspondences can translate to sound symbolic associations. The issue is that while crossmodal correspondences involve simple, unidimensional stimuli (e.g., pure tones), linguistic stimuli are by their very nature more perceptually complex and multidimensional. Thus, while it might be tempting (and potentially correct) to explain sound symbolic associations as arising from a crossmodal correspondence of a phoneme's component feature (e.g., between pitch and size; see Fig. 5), the fact that the feature is embedded in a multidimensional stimulus will necessarily complicate matters (see Parise, 2016). As an illustration of these complexities, D'Onofrio (2013) found that the influence of voicing in the maluma/takete effect was moderated by place of articulation. Nevertheless there is evidence that the two classes of effects are related. For instance, when Parise and Spence (2012) studied the mil/mal and maluma/takete effects using an IAT, they also examined crossmodal correspondences (e.g., between pitch and size). All of these effects were found to have the same effect size, which was interpreted as being indicative of a common mechanism.

A related question is the extent to which stimulus features are processed differently when they occur in linguistic versus nonlinguistic stimuli. For instance, does vowel pitch truly have the same associations as the pitch of a pure tone? Or does the involvement of the linguistic system alter the way that these associations operate (in addition to previously mentioned issues of multidimensionality)? Tsur (2006) theorized

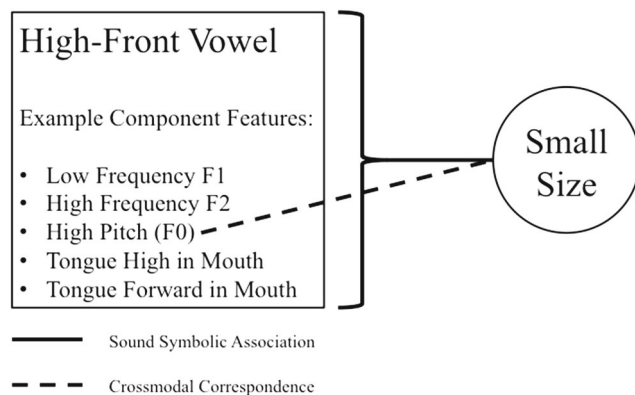


Fig. 5 The relationship between sound symbolic associations and crossmodal correspondences (as the terms are used in this review). Sound symbolic associations are between a phoneme as a whole (including all of its component multidimensional features) and a particular stimulus dimension. Crossmodal correspondences are between simple stimulus dimensions (e.g., brightness or pitch). Crossmodal correspondences may exist between the component features of a phoneme and a particular dimension (illustrated by the example crossmodal correspondence between phoneme pitch and size), and could potentially contribute to a sound symbolic association of that phoneme

that linguistic stimuli could be processed based on their phonetic identity, their sensory features, or a combination of the two. It would stand to reason that an overlap between sound symbolic associations and crossmodal correspondences would depend on the stimuli being processed (at least in part) based on their sensory features. Indeed there is some evidence for this. Fischer-Jørgensen (1968) found that Danish-speaking participants rated several pairs of allophones (e.g., [æ] and [a], allophones of /æ/, as in *had*), differently on semantic differential scales. While allophones belong to the same phoneme category, they have different sensory features. Thus the fact that they were rated differently indicates that their sensory features affected their interpretation. We would not have expected this if they had been processed solely in terms of their phonetic identity.

Next steps in exploring mechanisms of association

Resolving the issues above will add to our understanding of sound symbolism. Still remaining, however, is the question of which of the proposed mechanisms underlie these associations. It is our opinion that the current body of experimental evidence does not allow us to definitively pinpoint any particular mechanism(s) as being responsible for sound symbolism. This is in part because much of the existing work has focused on the effects of these associations, as opposed to the mechanisms that create them. Thus while a wealth of research exists, these experiments have not been designed to adjudicate between mechanisms of association.

It would be infeasible to test all five of the mechanisms at once, and so the best first step is likely to focus on developing testable hypotheses to adjudicate between two of them at a time. In our opinion, this initial pair of mechanisms should be statistical co-occurrence and shared properties (particularly connotations) since these two mechanisms are best supported by the available empirical evidence. There is compelling evidence that statistical co-occurrences can create crossmodal correspondences between dimensions (e.g., Baier, Kleinschmidt, & Müller, 2006; Ernst, 2007; Teramoto, Hidaka, & Sugita, 2010; Zangenehpour & Zatorre, 2010), though this still remains to be demonstrated for sound symbolic associations. There is also evidence of similar connotations creating crossmodal correspondences (e.g., L. Walker et al., 2012; P. Walker, 2012; P. Walker & Walker, 2012), and, more importantly, sound symbolic associations (e.g., Bozzi & Flores D'Arcais, 1967; French, 1977; Gallace et al., 2011). These results suggest these two mechanisms are the most promising starting points.

Using this pair of mechanisms as an example, future research aimed at adjudicating between mechanisms could take two tracks. One track would involve studies investigating whether each mechanism *can* create sound symbolic associations. This could be accomplished by attempting to create associations via either mechanism (e.g., artificially creating a

statistical co-occurrence, and associating unrelated stimulus dimensions with some unifying shared property). One might also hypothesize as yet unmeasured sound symbolic effects based on either mechanism, and then test for those novel effects.

The other track would involve studies that examine existing sound symbolic associations in terms of whether they are better explained by statistical co-occurrence or a shared property. If a particular effect is due to a statistical co-occurrence, then it should be possible to find evidence of that co-occurrence in the environment. In addition, we might expect manipulations of individuals' internalized probabilities to interfere with the association. If a particular effect is due to shared properties, then it should be possible to detect those shared properties via rating scales or reaction time measures. Another approach could be to examine if the strength of a given effect correlates with individual differences that would be relevant to a particular mechanism (e.g., differences in statistical learning; Misyak & Christiansen, 2012). Relatedly, sound symbolism effects have been shown to vary in some special populations (e.g., in individuals with autism spectrum disorder: Oberman & Ramachandran, 2008; Occelli, Esposito, Venuti, Arduino, & Zampini, 2013; in individuals with dyslexia: Drijvers, Zaadnoordijk, & Dingemans, 2015). To the extent that a given special population would be expected to differ in their capacity for a particular mechanism, this may represent another way of adjudicating between mechanisms.

Research that pits mechanisms against each other will be useful for generating evidence that some play a role in sound symbolism while others do not. While it is in principle possible that such research will discover that a single mechanism underlies all of sound symbolism, it seems more likely that multiple mechanisms contribute. The research reviewed in the preceding sections provides good reason to believe that a handful of mechanisms—even perhaps all those reviewed—play *some* role in sound symbolism. To the extent that this is borne out by future research, the next task for the field will be to examine the interplay between these mechanisms.

One possibility is that different mechanisms underlie different instances of sound symbolism. This suggests the intriguing possibility that certain mechanisms may be more likely to play a role for some kinds of dimensions than others. One potentially important distinction is that between prothetic (i.e., based on quantitative distinctions) and metathetic (i.e., based on qualitative distinctions) dimensions (Stevens, 1957). Gallace et al. (2011) hypothesized that for a metathetic domain such as taste, associations might be more likely to depend on shared *conceptual* properties. Conversely, an account such as magnitude coding requires a prothetic domain. Another relevant factor might be the salience and/or prevalence of a given stimulus dimension, which could potentially affect the likelihood of statistical co-occurrence playing a role. One might also expect evolutionary factors be more influential for

dimensions that are relevant to survival (e.g., size). Lastly, Ramachandran and Hubbard (2005) theorized that associations might be more likely to arise innately for stimulus dimensions that are represented in adjacent brain regions. Future research could compare mechanisms of association for dimensions that vary in these ways.

If it were demonstrated that different mechanisms underlie different effects, it would also be worthwhile for the field to consider if those different effects are indeed expressions of the same phenomenon. Perhaps it would be more accurate to view them as different kinds of sound symbolism—especially to the extent that they result in different behavioural effects. There is indeed some evidence of measurable differences between different instances of sound symbolism (e.g., Vainio et al., 2016). A distinction that is often made in the crossmodal correspondence literature is between perceptual and decisional effects. The former involve genuine differences in perception (e.g., perceiving a dot as moving upwards when presented along with rising pitch; Maeda, Kamai, & Shimojo, 2004) while the latter occur later in processing, and only involve effects on decisions, evident in reaction time or accuracy. Spence (2011) theorized that crossmodal correspondences arising from shared semantic features (in particular, shared labels) would not lead to perceptual effects, while those based on co-occurrences or neural factors *would* lead to perceptual effects. Investigating the perceptual/decisional effect distinction across instances of sound symbolism could be productive. We may also expect associations arising from some mechanisms not to emerge on implicit measures. For instance, as speculated, associations deriving from some shared properties may require explicit consideration. To the extent that associations with different origins lead to different behavioural outcomes, it may be prudent to consider them fundamentally different kinds of effects.

Another possibility is that multiple mechanisms combine to play a role in the same sound symbolic effect. For instance, it may be that the co-occurrence of two kinds of stimuli contributes to them having similar connotations. As noted, explanations based on shared properties beg the question of how stimuli come to be associated with those shared properties, and perhaps statistical co-occurrence could provide the answer in some instances.²² Conversely, it is possible that similar stimuli tend to co-occur more often (this is the basis of theories using lexical co-occurrence as a way of measuring meaning; e.g., Landauer & Dumais, 1997). Magnitude coding represents another instance of mechanisms interacting (i.e., shared properties and neural factors). In this case, stimuli from different dimensions have the shared property of high (or low)

²² However, statistical co-occurrence would certainly not apply in every instance. As P. Walker and Walker (2012) point out, though small and bright objects share connotations, we would not expect smallness to co-occur with surface brightness.

magnitude, but the association fundamentally results from the neural coding of that property.

This interplay between mechanisms seems especially relevant to evolution-based theories. Consider the fact that Ohala's (1994) frequency code hypothesis involves an evolved sensitivity to a statistical co-occurrence. This presents the intriguing possibility that while some statistical co-occurrence based associations must be learned, others have become innate via evolutionary processes. Note, however, that Ohala (1994) concedes that some postnatal experience may be required in the formation of the frequency code. Thus, perhaps it would be more correct to say that there is an evolved *predisposition to acquire* associations based on certain statistical co-occurrences. In particular, this seems more likely to apply to co-occurrences that are based on fundamental physical laws rather than those that may vary locally. Similarly, evolved predispositions may play a role in some phonemes and stimuli sharing affective properties (Nielsen & Rendall, 2011, 2013). In our review, we have treated each of the five mechanisms as distinct, but there are many ways in which they could interact in the production of sound symbolism. Moreover, some mechanisms may be so interdependent that they cannot be understood in isolation (e.g., if shared properties were to arise via co-occurrence).

As the preceding examples illustrate, while multiple mechanisms may play a role in a single effect, they need not do so simultaneously. On the contrary, several mechanisms may play out sequentially in the creation of an effect. This could be true in terms of both ontogeny and phylogeny. In addition, when considering the contribution of multiple mechanisms to an observed behavioural effect, some may be more proximally related to that effect than others. As an illustration, consider an instance in which statistical co-occurrence leads to stimuli sharing a connotation; while both mechanisms would contribute to an observed behavioural effect, the stimuli sharing a connotation may do so more proximally. Of course, it is also possible that in some effects, phonemes are simultaneously associated with stimuli by multiple *separate* mechanisms of association that do not interact (see D'Onofrio, 2013; Nichols, 1971). A major challenge for the field going forward will be untangling these complex interactions.

Conclusion

Sound symbolism refers to an association between phonemes and particular kinds of stimuli. It provides a means by which language can be nonarbitrary, by facilitating iconic relationships between form and meaning. A variety of mechanisms have been proposed to explain how acoustic and articulatory properties of phonemes come to be associated with other stimuli. The associations may arise due to phoneme features and related stimuli co-occurring in the world. Another possibility

is that phoneme features and associated stimuli share a common property, be it perceptual, conceptual, affective or linguistic. The associations may also be due to structural properties of the brain, evolution, or patterns extracted from language. While there is a wealth of experimental evidence on the effects of sound symbolic associations, there has been much less work on the mechanisms that might create them. It is our hope that the preceding review will foster such investigations. We suggest that future investigations should be focused around the following points:

- (a) Examining whether each of the different mechanisms *can* and *do* contribute to sound symbolic associations, potentially beginning with further investigation into the mechanisms of statistical co-occurrence and shared properties.
- (b) If evidence suggests that different mechanisms underlie different associations, examining whether some mechanisms are more likely for particular kinds of dimensions than others, and if associations created by different mechanisms result in different behavioural effects.
- (c) If evidence suggests that multiple mechanisms contribute to a particular sound symbolic effect, examining the interplay of those contributions.

The study of sound symbolism reveals hidden dimensions of richness and meaning in language. For instance, Jorge-Luis Borges (1980) opined that “the English [word] *moon* has something slow, something that imposes on the voice a slowness that suits the moon” (p. 62). We might speculate that this arises from the association between nasal sonorants (e.g., /m/ and /n/) and back vowels (e.g., /u/), and slowness (Cuskley, 2013; Saji et al., 2013). Such sound symbolic associations illuminate the multimodal nature of human cognition. As interest in sound symbolism increases, the focus of future research must shift to understanding the mechanisms that underlie such associations. The field must test predictions derived from extant theories, and work to refine those theories. We have offered some ideas for that future work here, and are confident that the years to come will bring with them a fuller and deeper understanding of this fascinating phenomenon.

Acknowledgements This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a postgraduate scholarship to D.M.S. and a Discovery Grant to P.M.P., and by Alberta Innovates: Health Solutions (AIHS) through a graduate scholarship to D.M.S.

We would like to thank our two reviewers for their helpful suggestions; Suzanne Curtin for her helpful comments on an earlier version of this manuscript; Michele Wellsby and Lenka Zdravilova for their helpful comments on a draft of this manuscript; Alberto Umiltà for providing translation of an article; and Padraic Monaghan for providing systematicity values.

Appendix A

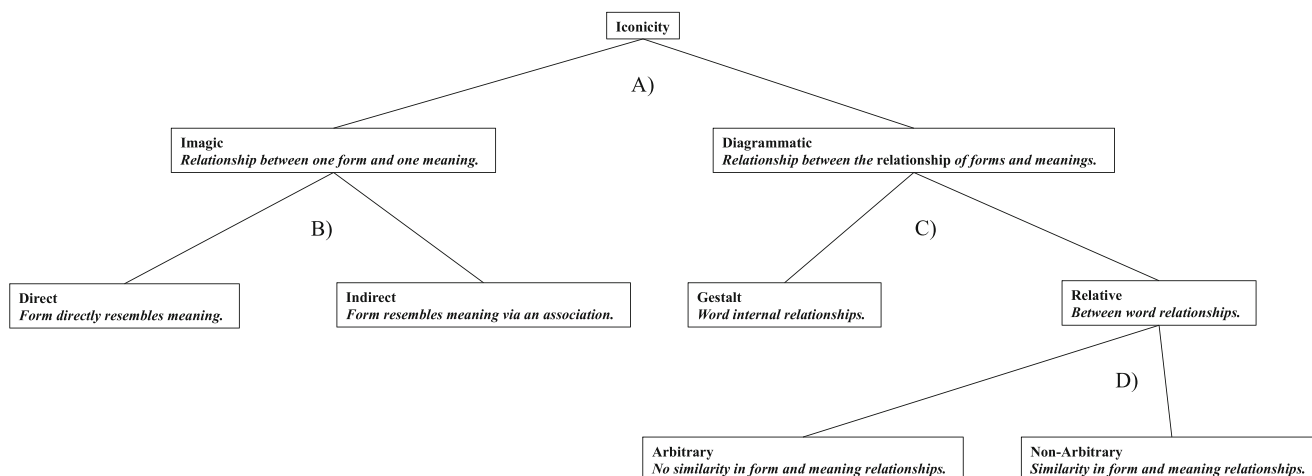


Fig. 6 A proposed taxonomy of iconicity. In all cases of iconicity there is a structural resemblance-based mapping between form and meaning; however, this can happen in different ways. The figure is an attempt to integrate the various distinctions that have been made in the literature. **a** In imagic iconicity, the iconic relationship is between a single form and its meaning; in diagrammatic iconicity, the iconic relationship is between the relationships among forms and the relationships among their meanings (i.e., a relationship between relationships; Peirce, 1974). Note that this distinction is sometimes referred to as being between *absolute* and *relative* iconicity (e.g., Dingemanse et al., 2015). **b** In direct imagic iconicity a form directly resembles its meaning (e.g., onomatopoeia); in indirect imagic iconicity a form resembles its meaning by way of its associations (e.g., words with sound symbolically small phonemes

[e.g., /i/] with small meanings; Masuda, 2007)^a. **c** In gestalt diagrammatic iconicity, the form relationship is within a single word (e.g., word structure iconically conveying event structure; for instance, closed syllables evoking end points); in relative diagrammatic iconicity the form relationship is between two (or more) words (Dingemanse, 2012). **d** Arbitrary relative iconicity refers to the mere existence of similar forms with similar meanings (e.g., phonesthemes); nonarbitrary relative iconicity refers to instances in which the relationship among word forms is similar to the relationship among their meanings (e.g., three words with progressively higher vowels, meaning progressively smaller things). ^aWhile we attribute the term *indirect iconicity*—meaning sound symbolically mediated iconicity—to Masuda (2007), that author did not use it specifically in the context of imagic iconicity as we have here

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson, S. R. (1985). *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago: The University of Chicago Press.
- Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in pre-verbal infants. *Cortex*, 63, 196–205. doi:10.1016/j.cortex.2014.08.025
- Assaneo, M. F., Nichols, J. I., & Trevisan, M. A. (2011). The anatomy of onomatopoeia. *PLOS ONE*, 6. doi:10.1371/journal.pone.0028317
- Baier, B., Kleinschmidt, A., & Müller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *Journal of Neuroscience*, 26, 12260–12265. doi:10.1523/JNEUROSCI.1457-06.2006
- Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, 136, 186–195. doi:10.1016/j.cognition.2014.11.013
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, 17, 645–648. doi:10.1111/j.1467-9280.2006.01759.x
- Bentley, M., & Varon, E. J. (1933). An accessory study of "phonetic symbolism". *The American Journal of Psychology*, 45, 76–86. doi:10.2307/1414187
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80, 290–311. doi:10.1353/lan.2004.0056
- Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 76–93). Cambridge, UK: Cambridge University Press.
- Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Timonium: York Press.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National*

- Academy of Sciences*, 113, 10818–10823. doi:10.1073/pnas.1605782113
- Bolinger, D. (1950). Rime, assonance, and morpheme analysis. *Word*, 6, 117–136. doi:10.1080/00437956.1950.11659374
- Borges, J. L. (1980). *Seven nights*. New York: New Directions.
- Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K. H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, 7. doi:10.1038/srep41070
- Bozzi, P., & Flores D'Arcais, G. B. (1967). Experimental research on the intermodal relationships between expressive qualities. *Archivio di Psicologia, Neurologia e Psichiatria*, 28, 377–420.
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). “bouba” and “kiki” in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, 126, 165–172. doi:10.1016/j.cognition.2012.09.007
- Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology*, 50, 388–393.
- Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30, 348–369. doi:10.1016/0749-596X(91)90041-H
- Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, 7, 12739. doi:10.1038/ncomms12739
- Cowles, J. T. (1935). An experimental study of the pairing of certain auditory and visual stimuli. *Journal of Experimental Psychology*, 18, 461–469.
- Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5, 39–62.
- Cuskley, C., Simner, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research*. doi:10.1007/s00426-015-0709-2
- D'Onofrio, A. (2013). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech*, 57, 367–393. doi:10.1177/0023830913507694
- de Saussure, F. (1916). *Course in General Linguistics*. New York: Columbia University Press.
- Davies, N. B., & Halliday, T. R. (1978). Deep croaks and fighting assessment in toads *Bufo bufo*. *Nature*, 274, 683–685. doi:10.1038/274683a0
- Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52, 259–268. doi:10.1111/j.2044-8295.1961.tb00788.x
- Deroy, O., & Auvray, M. (2013). A new Molyneux's problem: Sounds, shapes and arbitrary crossmodal correspondences. In O. Kutz, M. Bhatt, S. Borgo, & P. Santos (Eds.), *Second International Workshop on the Shape of Things* (pp. 61–70).
- Deroy, O., Crisinel, A. S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review*, 20, 878–896. doi:10.3758/s13423-013-0397-0
- Diffolth, G. (1994). i: big, a: small. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 107–114). Cambridge: Cambridge University Press.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Linguistics and Language Compass*, 6, 654–672. doi:10.1002/lnc3.361
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19, 603–615. doi: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, e67–e83. doi:10.1353/lan.2016.0034
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, 24, 613–621. doi:10.1177/0956797612457374
- Drijvers, L., Zaadnoordijk, L., & Dingemanse, M. (2015). Sound-symbolism is disrupted in dyslexia: Implications for the role of cross-modal abstraction processes. In D. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015)* (pp. 602–607). Cognitive Science Society, Austin.
- Eberhardt, M. (1940). A study of phonetic symbolism of deaf children. *Psychological Monographs*, 52, 23–41.
- Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369. doi:10.1098/rstb.2013.0301
- Ernst, O. M. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7, 1–14. doi:10.1167/7.5.7
- Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology*, 72, 160–163. doi:10.1016/j.biopsycho.2005.09.003
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15, 399–402. doi:10.1046/j.0953-816x.2001.01874.x
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 12203–12208. doi:10.1073/pnas.0602173103
- Feld, S. (1982). *Sound and sentiment: Birds, weeping, poetics, and song in Kaluli expression*. Philadelphia: University of Pennsylvania Press.
- Fischer-Jørgensen, E. (1968). Perceptual dimensions of vowels. *STUF-Linguage Typology and Universals*, 21, 94–98.
- Fischer-Jørgensen, E. (1978). On the universal character of phonetic symbolism with special reference to vowels. *Studia Linguistica*, 32, 80–90.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102, 1213–1222. doi:10.1121/1.421048
- Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36, 967–997. doi:10.1017/S0305000908009252
- Fort, M., Martin, A., & Peperkamp, S. (2014). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech*, 58, 247–266. doi:10.1177/0023830914534951
- Fort, M., Weiß, A., Martin, A., & Peperkamp, S. (2013). *Looking for the bouba-kiki effect in pre-lexical infants*. Poster presented at the International Child Phonology Conference, Nijmegen, The Netherlands.
- French, P. L. (1977). Towards an explanation of phonetic symbolism. *Word*, 28, 305–322. doi:10.1080/00437956.1977.11435647
- Gallace, A., Boschini, E., & Spence, C. (2011). On the taste of “bouba” and “kiki”: An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience*, 2, 34–46. doi:10.1080/17588928.2010.516820
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, 68, 1191–1203. doi:10.3758/BF03193720

- Gamkrelidze, T. V. (1974). The problem of 'l'arbitraire du signe'. *Language*, 50, 102–110.
- Gasser, M. (2004). The origins of arbitrariness in language. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society*. 26; 434–439.
- Gentilucci, M., & Campione, G. C. (2011). Do postures of distal effectors affect the control of actions of other distal effectors? Evidence for a system of interactions between hand and mouth. *PLOS ONE*, 6. doi:10.1371/journal.pone.0019793
- Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, 30, 949–960. doi:10.1016/j.neubiorev.2006.02.004
- Gingras, B., Boeckle, M., Herbst, C. T., & Fitch, W. T. (2013). Call acoustics reflect body size across four clades of anurans. *Journal of Zoology*, 289, 143–150. doi:10.1111/j.1469-7998.2012.00973.x
- Gordon, M., & Heath, J. (1998). Sex, sound Symbolism, and sociolinguistics. *Current Anthropology*, 39, 421–449.
- Greenberg, J. H. (1978). Introduction. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of language, Volume 2: Phonology* (pp. 1–8). Redwood City, CA: Stanford University Press.
- Greenberg, J. H., & Jenkins, J. J. (1966). Studies in the psychological correlates of the sound system of American English. *Word*, 22, 207–242.
- Hauser, M. D. (1993). The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context. *The American Naturalist*, 142, 528–542.
- Hinton, L., Nichols, J., & Ohala, J. J. (1994). Sound-symbolic processes. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 1–14). Cambridge, UK: Cambridge University Press.
- Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner's speeded classification task. *Perceptual and Motor Skills*, 113, 929–940. doi:10.2466/15.21.28.PMS.113.6.929-940
- Hockett, C. (1963). The problem of universals in language. In J. Greenberg (Ed.), *Universals of language* (pp. 1–22). Cambridge, MA: MIT Press.
- Hung, S.-M., Styles, S. J., & Hsieh, P.-J. (2017). Can a word sound like a sharp before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science*, 28, 263–275. doi:10.1177/0956797616677313
- Hutton, C. (1989). The arbitrary nature of the sign. *Semiotica*, 75, 63–78. doi:10.1515/semi.1989.75.1-2.63
- Ikegami, T., & Zlatev, J. (2007). From non-representational cognition to language. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, language and mind, Vol 1: Embodiment* (pp. 241–283). Berlin: Mouton.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369. doi:10.1098/rstb.2013.0298
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54–65. doi:10.1016/j.cognition.2008.07.015
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13, 21–37.
- Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
- Jakobson, R., & Waugh, L. (1979). *The sound shape of language*. Bloomington: Indiana University Press.
- Jespersen, O. (1922). The symbolic value of the vowel *i*. *Philologica*, 1, 1–19.
- Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *The Public Journal of Semiotics*, 5, 3–20.
- Johnson, R. C., Suzuki, N. S., & Olds, W. K. (1964). Phonetic symbolism in an artificial language. *The Journal of Abnormal and Social Psychology*, 69, 233–236. doi:10.1037/h0043851
- Joseph, J. E. (2015). Iconicity in Saussure's linguistic work, and why it does not contradict the arbitrariness of the sign. *Historiographia Linguistica*, 42, 85–105.
- Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLOS ONE*, 9, 1–8. doi:10.1371/journal.pone.0097905
- Kantartzis, K. F. (2011). *Children and adults' understanding and use of sound-symbolism in novel words* (Doctoral dissertation). Retrieved from eTheses Repository (2997).
- Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. *The role of form in visual responses to music*. *Journal of General Psychology*, 26, 199–222.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 419–454
- Kirkham, N. Z., Slemner, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42. doi:10.1016/S0010-0277(02)00004-5
- Klank, L. J., 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, 140–148. doi:10.1016/S0022-5371(71)80005-1
- Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.
- Komeilipoor, N., Tiainen, M., Tiippana, K., Vainio, M., & Vainio, L. (2016). Excitability of hand motor areas during articulation of syllables. *Neuroscience Letters*, 620, 154–158. doi:10.1016/j.neulet.2016.04.004
- Koriat, A., & Levy, I. (1977). The symbolic implications of vowels and of their orthographic representations in two natural languages. *Journal of Psycholinguistic Research*, 6, 93–103. doi:10.1007/BF01074374
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114, 19–28. doi:10.1016/j.cognition.2009.08.016
- Ladefoged, P., & Johnson, K. (2010). *A course in linguistics* (6th ed.). Boston: Wadsworth.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240. doi:10.1037/0033-295X.104.2.211
- Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, 16, 597–607. doi:10.1037/0012-1649.16.6.597
- Lockwood, G., & Dingemans, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.01246
- Lockwood, G., Dingemans, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1274–1281.
- Lockwood, G., Hagoort, P., & Dingemans, M. (2016). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra*, 2, 1–15. doi:10.1525/collabra.42
- Lockwood, G., & Tuomainen, J. (2015). Ideophones in Japanese modulate the P2 and late positive complex responses. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.00933
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences*, 108, 20661–20665.

- Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, 7, 167–193. doi:10.1017/langcog.2014.21
- Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, 14, R990–R991. doi:10.1016/j.cub.2004.11.018
- Magnus, M. (2000). *What's in a word? Evidence for phonosemantics* (Doctoral dissertation). Retrieved from NTNU Open (82-471-5073-5).
- Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, 87, 173–188.
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 384–394. doi:10.1037/0096-1523.13.3.384
- Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 586–602. doi:10.1037/0096-1523.15.3.586
- Marks, L. E. (2013). Weak synesthesia in perception and language. In J. Simner, & E. H. Hubbard (Eds.), *The Oxford handbook of synesthesia* (pp. 761–789). Oxford: Oxford University Press.
- Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, 28, 903–923. doi:10.1068/p2866
- Masuda, K. (2007). The physical basis for phonological iconicity. In E. Tabakowska, C. Ljungberg, & O. Fischer (Eds.), *Insistent images* (pp. 57–72). Philadelphia: John Benjamins.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316–322. doi:10.1111/j.1467-7687.2006.00495.x
- Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2013). Competing iconicities in the structure of languages. *Cognitive Linguistics*, 24, 309–343. doi:10.1515/cog-2013-0010
- Miron, M. S. (1961). A crosslinguistic investigation of phonetic symbolism. *The Journal of Abnormal and Social Psychology*, 62, 623–630. doi:10.1037/h0045212
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302–331. doi:10.1111/j.1467-9922.2010.00626.x
- Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140, 325–347. doi:10.1037/a0022924
- Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1152–1164. doi:10.1037/a0027747
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369. doi:10.1098/rstb.2013.0299
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855–869.
- Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge, UK: Cambridge University Press.
- Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, 45, 53–75. doi:10.2307/1414186
- Nichols, J. (1971). Diminutive consonant symbolism in western North America. *Language*, 47, 826–848. doi:10.2307/412159
- Nielsen, A. K. S., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65, 115–124. doi:10.1037/a0022268
- Nielsen, A. K. S., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 67, 153–163. doi:10.1037/a0030553
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28, 225–252. doi:10.1146/annurev.anthro.28.1.225
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112, 181–186. doi:10.1016/j.cognition.2009.04.001
- Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience*, 3, 348–355. doi:10.1080/17470910701563681
- Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The Takete-Maluma phenomenon in autism spectrum disorders. *Perception*, 42, 233–241. doi:10.1068/p7357
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge, UK: Cambridge University Press.
- Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Eds.), *In honour of Ilse Lehiste* (pp. 207–215). Dordrecht: Foris.
- Ohtake, Y., & Haryu, E. (2013). Investigation of the process underpinning vowel-size correspondence. *Japanese Psychological Research*, 55, 390–399. doi:10.1111/jpr.12029
- Osgood, C. E., Suci, S. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews*, 10, 58–71. doi:10.1002/evan.1014
- 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, 173–186. doi:10.1016/j.jecp.2012.05.004
- Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29, 7–28. doi:10.1163/22134808-00002502
- Parise, C. V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, 214, 373–380. doi:10.1007/s00221-011-2836-3
- Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, 220, 319–333. doi:10.1007/s00221-012-3140-6
- Parise, C. V., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In J. Simner & E. Hubbard (Eds.), *Oxford handbook of synaesthesia* (pp. 790–815). Oxford: Oxford University Press.
- Patel, R., Mulder, R. A., & Cardoso, G. C. (2010). What makes vocalisation frequency an unreliable signal of body size in birds? A study on black swans. *Ethology*, 116, 554–563. doi:10.1111/j.1439-0310.2010.01769.x
- Peirce, C. S. (1974). *Collected papers of Charles Sanders Peirce* (6th ed.). Boston: Harvard University Press.
- Pejovic, J., & Molnar, M. (2016). The development of spontaneous sound-shape matching in monolingual and bilingual infants during the first year. *Developmental Psychology*, 53, 581–586. doi:10.1037/dev0000237
- Peña, M., Mehler, J., & Nespors, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, 22, 1419–1421. doi:10.1177/0956797611421791

- Permiss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*. doi:10.3389/fpsyg.2010.00227
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLOS ONE, 10*. doi:10.1371/journal.pone.0137147
- Preziosi, M. A., & Coane, J. H. (2017). Remembering that big things sound big: Sound symbolism and associative memory. *Cognitive Research: Principles and Implications, 2*. doi:10.1186/s41235-016-0047-y
- Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., & Trope, Y. (2016). The sound of distance. *Cognition, 152*, 141–149. doi:10.1016/j.cognition.2016.04.001
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3–34.
- Ramachandran, V. S., & Hubbard, E. M. (2005). The emergence of the human mind: Some clues from synesthesia. In L. C. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 147–192). Oxford: Oxford University Press.
- Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, production, acoustics, and perception*. Hoboken: Wiley-Blackwell.
- Reichard, G. A. (1944). *Prayer: The compulsive word (American Ethnological Society Monograph, 7)*. Seattle: University of Washington Press.
- Reichard, G. A. (1950). *Navaho religion: A study of symbolism*. New York: Pantheon Books.
- Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science, 31*, 157–168. doi:10.1080/03640210709336988
- Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America, 113*, 3390–3402. doi:10.1121/1.1568942
- Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America, 117*, 944–955. doi:10.1121/1.1848011
- Rogers, S. K., & Ross, A. S. (1975). A cross-cultural test of the Maluma-Takete phenomenon. *Perception, 4*, 105–106.
- Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion, 14*, 246–250. doi:10.1037/a0035752
- Saji, N., Akita, K., Imai, M., Kantartzis, K., & Kita, S. (2013). Cross-linguistically shared and language-specific sound symbolism for motion: An exploratory data mining approach. *Proceedings of the 35th Annual Conf. of the Cognitive Science Society, 31*, 1253–1259.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225–239.
- Sedley, D. (2013). Plato's *Cratylus*. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2013 ed.). Retrieved from <https://plato.stanford.edu/entries/plato-cratylus/>
- Sereno, J. A. (1986). Stress pattern differentiation of form class in English. *The Journal of the Acoustical Society of America, 79*, S36. doi:10.1121/1.2023191
- Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society, 6*, 96–105. doi:10.2752/174589311X12893982233911
- Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Proceedings of the 36th Annual Meeting of the Berkeley Linguistics Society*. doi:10.3765/bls.v36i1.3926
- Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language, 55*, 167–177. doi:10.1016/j.jml.2006.03.002
- Sidhu, D. M., & Pexman, P. M. (2015). What's in a name? Sound symbolism and gender in first names. *PLOS ONE, 10*. doi:10.1371/journal.pone.0126809
- Sidhu, D. M., & Pexman, P. M. (2016). A prime example of the Maluma/Takete effect? Testing for sound symbolic priming. *Cognitive Science*. doi:10.1111/cogs.12438
- Sidhu, D. M., & Pexman, P. M. (2017). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2017.1358379
- Sidhu, D. M., Pexman, P. M., & Saint-Aubin, J. (2016). From the Bob/Kirk effect to the Benoit/Éric effect: Testing the mechanism of name sound symbolism in two languages. *Acta Psychologica, 169*, 88–99. doi:10.1016/j.actpsy.2016.05.011
- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24*, 99–142. doi:10.1016/0010-0285(92)90004-L
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*, 971–995. doi:10.3758/s13414-010-0073-7
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181. doi:10.1037/h0046162
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54*, 768–777. doi:10.1037/0022-3514.54.5.768
- Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape-sound-meaning congruence in words, and shape-sound congruence in pseudowords. *Brain and Language, 145*, 11–22. doi:10.1016/j.bandl.2015.03.011
- Sullivan, B. K. (1984). Advertisement call variation and observations on breeding behavior of *Bufo debilis* and *B. punctatus*. *Journal of Herpetology, 18*, 406–411. doi:10.2307/1564103
- Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech, 14*, 266–276. doi:10.1177/002383097101400307
- Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 21*, 352–360. doi:10.1016/S0022-5371(82)90670-3
- Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge, UK: Cambridge University Press.
- Taylor, I. K. (1963). Phonetic symbolism revisited. *Psychological Bulletin, 60*, 200–209.
- Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds move a static visual object. *PLOS ONE, 5*. doi:10.1371/journal.pone.0012255
- Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology, 64*, 37–41. doi:10.1080/17470218.2011.605898
- Tsur, R. (2006). Size-sound symbolism revisited. *Journal of Pragmatics, 38*, 905–924. doi:10.1016/j.pragma.2005.12.002
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*, 769–800. doi:10.1080/13506280042000144
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*, 4–21. doi:10.1159/000356237
- Urban, M. (2011). Conventional sound symbolism in terms for organs of speech: A cross-linguistic study. *Folia Linguistica, 45*, 199–214. doi:10.1515/flin.2011.007

- Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. (2013). Effect of Syllable articulation on precision and power grip performance. *PLOS ONE*, 8. doi:10.1371/journal.pone.0053061
- Ullian, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language*. Vol. 2: Phonology (pp. 525–568). Stanford: Stanford University Press.
- Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M. (2016). Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. *Psychological Research*, 81, 827–839. doi: 10.1007/s00426-016-0778-x
- Van Lancker, D., & Cummings, J. L. (1999). Expletives: Neurolinguistic and neurobehavioral perspectives on swearing. *Brain Research Reviews*, 31, 83–104. doi:10.1016/S0165-0173(99)00060-0
- Velasco, C., Woods, A. T., Deroy, O., & Spence, C. (2015). Hedonic mediation of the crossmodal correspondence between taste and shape. *Food Quality and Preference*, 41, 151–158. doi:10.1016/j.foodqual.2014.11.010
- von der Gabelentz, G. (1891). *Die sprachwissenschaft: Ihre aufgaben, methoden und bisherigen ergebnisse* [Linguistics: Its functions, methods and results so far]. Leipzig: T. O. Weigel.
- Von Humboldt, W. (1836). *On language: On the diversity of human language construction and its influence on the mental development of the human species*. Cambridge, UK: Cambridge University Press.
- Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., ... & Bulnes, L. C. (2016). Registered replication report Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, 11, 917–928. doi:10.1177/1745691616674458
- Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics*, 74, 1792–1809. doi:10.3758/s13414-012-0341-9
- Walker, P. (2016). Cross-sensory correspondences and symbolism in spoken and written language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1339–1361. doi:10.1037/xlm0000253
- Walker, P., & Walker, L. (2012). Size–brightness correspondence: Crosstalk and congruity among dimensions of connotative meaning. *Attention, Perception, & Psychophysics*, 74, 1226–1240. doi:10.3758/s13414-012-0297-9
- Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception*, 41, 1186–1192. doi:10.1068/p7149
- Wallschläger, D. (1980). Correlation of song frequency and body weight in passerine birds. *Cellular and Molecular Life Sciences*, 36, 412. doi:10.1007/BF01975119
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quality. *Trends in Cognitive Sciences*, 7, 483–488. doi:10.1016/j.tics.2003.09.002
- Wan, X., Woods, A. T., van den Bosch, J. J. F., McKenzie, K. J., Velasco, C., & Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between basic tastes and visual features. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.01365
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989–994. doi:10.1016/S0028-3932(02)00316-0
- Waugh, L. R. (1992). Presidential address: let's take the con out of iconicity. *American Journal of Semiotics*, 9, 7–47.
- Waugh, L. R. (1993). Against arbitrariness: Imitation and Motivation revived, with consequences for textual meaning. *Diacritics*, 23, 71–87. doi:10.2307/465317
- Werner, H., & Kaplan, B. (1963). *Symbol formation. An organismic-developmental approach to language and the expression of thought*. New York: Wiley.
- Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, 93, 10–19. doi:10.1016/j.bandl.2004.07.006
- Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2017). *Weighing up the evidence for sound symbolism: Distributional properties predict cue strength*. Manuscript submitted for publication.
- Westermann, D. H. (1927). Laut, Ton und Sinn in westafrikanischen Sudansprachen [Sound, tone and meaning in the West African languages of Sudan]. In F. Boas (Ed.), *Festschrift Meinhof*. Hamburg: L. Friederichsen.
- Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy*, 15, 844–858. doi:10.3390/e12040844
- Winter, B., Perlman, M., Perry, L. K. & Lupyan, G. (2017). Which words are the most iconic? Iconicity in English sensory words. *Interaction Studies*. doi:10.1075/is.18.3.07win
- Witherspoon, G. (1977). *Language and art in the Navajo universe*. Ann Arbor: University of Michigan Press.
- Wu, L., Klink, R. R., & Guo, J. (2013). Creating gender brand personality with brand names: The effects of phonetic symbolism. *Journal of Marketing Theory and Practice*, 21, 319–330. doi:10.2753/MTP1069-6679210306
- Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Review*, 96, 395–416. doi:10.1037/0033-295X.96.3.395
- Zangenehpour, S., & Zatorre, R. J. (2010). Cross-modal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, 48, 591–600. doi:10.1016/j.neuropsychologia.2009.10.022