Forms, Mechanisms, and Roles of Iconicity in Spoken Language: A Review

Harry Barker,[1] Mirjana Bozic[1]

[1] *Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, UK*

Abstract

Historically, debates over relationships between spoken lexical form and meaning have been dominated by views of arbitrariness. However more recent research revealed a different perspective, in which non-arbitrary mappings play an important role in the makeup of a lexicon. It is now clear that phoneme-sound symbolism - along with other types of form-to-meaning mappings - contributes to non-arbitrariness (iconicity) of spoken words, which is present in many forms and degrees in different languages. Attempts have been made to provide a mechanistic explanation of the phenomenon, and these theories largely centre around cross-modal correspondences. We build on these views to explore iconicity within the evolutionary context and the neurobiological framework for human language processing. We argue that the multimodal bihemsipheric communicative system, to which iconicity is integral, has important phylogenetic and ontogenetic advantages, facilitating language learning, comprehension, and processing. Despite its numerous advantages however, iconicity must compete with arbitrariness, forcing language systems to balance the competing needs of perceptual grounding of the linguistic form and ensuring an effective signal. We conclude that, on balance, iconicity should be viewed as integral to language, and not merely a marginal phenomenon.

Keywords

Iconicity, sound symbolism, cross-modal correspondence, language evolution, neurobiology of language

1. Introduction

The nature of the relationship between word forms and their meanings has long been a prominent question. Do words sound like what they describe, or is this relationship arbitrary? Modern linguistic theory has largely adopted the view that the phonological form of the lexical sign bears no resemblance to its meaning; the connection is arbitrary. This view is often attributed to Saussure (1959) who argued "Because the sign is arbitrary, it follows no law other than that of tradition, and because it is based on tradition, it is arbitrary". However, the idea of arbitrariness was established much earlier by Locke (1690), who proposed that the existence of multiple languages is sufficient evidence to disprove a so-called 'natural' connection: if the properties of the concept or object that the word refers to are sufficient to determine a phonological form, then there should only be one language. The notion of arbitrariness provides a robust explanation for the origin of most words; for example, there is nothing about the phonological form of the word 'pen' to suggest that it should denote 'an instrument for writing or drawing with ink'. It has simply been established that there is a conventionalised system of symbols shared by communities of users and passed from generation to generation - "there is no reason for you to call a dog 'dog' rather than 'cat' except for the fact that everyone else is doing it" (Pinker & Bloom, 1990, p.728). Notable exceptions to the conventional approach, such as onomatopoeia, have been regarded as marginal phenomena within an arbitrary linguistic system, with Saussure (1959) arguing that "such words are never organic elements" of language. In addition, although onomatopoeic words seem to represent the sounds they symbolise, they too show evidence of arbitrariness. For example, the English expression for the sound of a rooster (*cock-a-doodle-doo*) differs from German and French expressions (*kikeriki* and *cocorico*, respectively), all of which are acoustically different from the actual sound made by a rooster (Perniss et al., 2010). There is therefore plenty of evidence to support the idea of conventionalised, wholly arbitrary mapping between phonological form and meaning in our communication system. Indeed, the idea of language as arbitrary dominates many of the prominent theories of language acquisition, production, and comprehension; where relatively little

attention has been focused on determining how phonological and semantic representations might overlap.

However, in looking for an accurate account of non-arbitrariness in language, a broad-cross linguistic perspective is required, and more recent work has looked beyond Indo-European languages as well as at both signed and spoken languages (Perlman et al., 2018). This research has suggested a more textured view that non-arbitrariness may have a more important role in language than previously thought, and may in fact complement arbitrariness by conveying different linguistic advantages.

1.1 Defining Forms of Non-arbitrariness

When considering forms of non-arbitrariness in language, there has been much confusion over terminology despite attempts to clarify these uncertainties (e.g. Sidhu, 2019; Lockwood and Dingemanse, 2015). The term *iconicity* is usually taken to refer to a form-meaning resemblance that may apply to spoken, manual and written communication and is not limited to human language. In fact, there is plenty of evidence for iconicity in the communicative acts of non-human primates, which we will later explore in the context of human language evolution. *Sound symbolism,* which the current paper focuses on, is used to specifically refer to iconicity in spoken language, though these terms have previously been used interchangeably in much of the literature.

It has been argued that iconicity is not a binary property and may be present in different degrees, with a distinction between 'absolute' (imagic) and 'relative' (diagrammatic) iconicity (Dingemanse et al., 2015; Peirce, 1974); and 'direct' and 'indirect' iconicity (Masuda, 2007). Absolute iconicity is where there is a one-to-one mapping between phonological form and meaning. This may be 'direct' as in onomatopoeia, for example the sound of water leaking '*drip drop'* (English) or '*plitsch platsch'* (German), or 'indirect', such as the association between the phoneme /i/ and smallness in words like '*petite'* and '*tiny'*. Absolute iconicity in spoken language has also been labelled *sensory sound symbolism* (Cuskley and Kirby, 2013), in other words the phenomenon where a word's form imitates its referent (Lockwood and Dingemanse, 2015). In

relative iconicity, (what Cuskley and Kirby, (2013) refer to as *conventional sound symbolism*) the relationships between multiple forms are analogous to relations between different meanings. This can be thought of as a mapping between two or more words and two or more referents, such as phonesthemes like the English 'str-', which denotes something long and thin, as in '*straight'*, '*stripe'*, '*stream'*, and '*string'* (though it is worth noting that it is debated whether phonesthemes have a direct sensory link to associated meaning, as such 'pockets of consistency' in the lexicon may arise simply by chance (Baayen et al, 2011)). However some phonesthemes e.g. 'sn-' (occurring in words relating to the nose – '*sniff*', '*snore'*, '*sneeze*') do ilicit sensory imagery and could therefore be considered iconic, Schmidtke et al., (2014)). The term *conventional sound symbolism* can also cover correlations between sounds and grammatical categories – which is a form of non-arbitrariness called *systematicity* (Monaghan et al., 2014; Dingemanse et al., 2015). A good example is nouns denoting abstract concepts being longer and more derivationally complex than concrete nouns (Reilly & Kean, 2007). This is a form of non-arbitrariness as it reveals connections between word forms and meanings; however it primarily concerns broad statistical patterns between groups of words across the lexicon, rather than the mappings between specific words and their meanings (see Dingemanse et al., (2015) for a detailed review).

This paper will first briefly review evidence for some of the different forms of sound symbolism in human language, and build on existing work (Dingemanse et al. 2015; Sidhu, 2019; Sidhu & Pexman, 2018; Svantesson, 2017) to explore how these form-meaning relationships might arise. We will then turn our attention to the possible role of sound symbolism from an ontogenetic and evolutionary perspective. Finally, we incorporate evidence for iconicity in non-human communication within a neurobiological framework for human language processing to suggest that sound-symbolism, in addition to arbitrariness, should be viewed as integral to language, rather than a marginal linguistic phenomenon.

2. Sound Symbolism

Spoken languages consist of an inventory of different sounds, or phonemes, which can be combined in different ways to form words. These words are then attributed to meanings, leading to the fundamental questions about the nature of this form-meaning relationship raised above. Primarily, as a foundation for spoken language, do phonemes themselves convey inherent qualities which lead to association with certain meanings, or are they semantically neutral in nature? The notion that phonemes may carry such associations underlies the phenomenon of sound symbolism. There are different ways in which phonemes may evoke semantic properties, examples of which are presented below.

2.1. Sound-Size Symbolism

One of the best-known examples of sound symbolism is the idea that different vowels may communicate the semantic property of either small or large size. Almost a century ago it was first reported that a closed-front vowel /i/ may more naturally denote small size over an open-front vowel /a/ (Sapir, 1929). Sapir presented participants with short nonwords such as *'mil'* and *'mal'* that were allocated to meanings such as '*table'*. Participants were then asked to distinguish these nonwords according to their perceived size and decide which nonword symbolised a larger object. Between 75-96% of responses were in favour of the nonword containing /a/ as referring to the larger referent, consistently across different ages and language backgrounds (although with a large variance in individual sensitivity to the 'symbolic suggestiveness' of phoneme features). This was later extended to include a more gradual scale, as well as vowels in the front-back dimension (ɔ = o > u = a > æ > ɛ > e > i; Newman, 1933). Outside of forced-choice experiments, the association between different phonemes and the depiction of size has since been corroborated by cross-linguistic evidence, with some languages containing very elaborate size scales. One example is Khmu, the language of the northern Laos region. According to Svantesson (2017), in Khmu ideophones (words that represent sensory imagery), vowel variation is used to indicate size differences, as illustrated by the expressions for 'drink noisily' (*crúut-crúut* > *cróot-cróot* > *crüt-crüt* >

créet-créet). The expression chosen may vary according to the size of the drinking animal - for example, *crúut-crúut* could describe the sound of a buffalo drinking noisily, *cróot-cróot* the sound of a human, and *críit-críit/créet-créet* the sound of smaller animals (Figure 1a).

There is also evidence to suggest that it is the number of 'small' or 'large' vowel sounds in a word, and not vowel alone, that determines the conveyed magnitude (Thompson & Estes, 2011). In this study, English-speaking participants were presented with novel figures of varying size and asked to match them to nonwords containing either 'large' (*u, o*) or 'small' (*i, e*) vowels. The authors found a linear relationship between the size of an object and the number of 'large' phonemes in its preferred name, regardless of whether the nonwords were presented aurally or in writing.

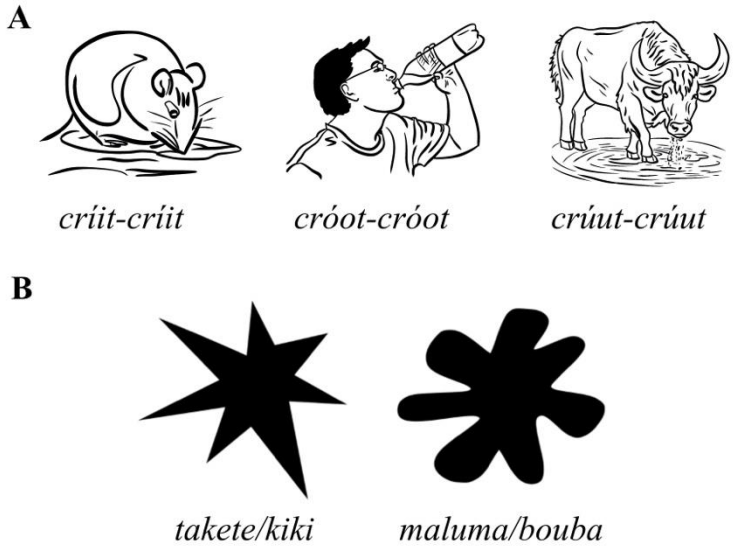


Figure. 1. a) The use of vowel variation to indicate size differences in Khmu ideophones (Svantesson; 2017). b) The takete/maluma or bouba/kiki effect - when prompted most subjects call the angular shape (left) *takete/kiki* and the rounded (right) shape *maluma/bouba*.

2.2. Sound-Shape Symbolism

Another well-documented example of sound symbolism is the '*maluma/takete effect'*, which denotes a link between specific phonemes and the visual perception of roundness or sharpness. Early evidence for this effect was again largely drawn from forced-choice experiments using non-words. Nearly a century ago, Köhler (1929) presented two shapes, one with more rounded edges and one with pointed edges (Figure 1b) and asked participants to choose which shape to assign

147 the names '*takete'* or '*maluma*' to. Ramachandran & Hubbard (2001) adapted the figures slightly and
148 changed the names to '*bouba'* and '*kiki*'. Both studies reported that *takete* / *kiki* label was more
149 likely to be asociated with the pointly shape, and *maluma* / *bouba* with the rounded shape; with
150 Ramachandran & Hubbard (2001) estimating a 95% prevalence of the expected response in
151 English-speaking adult participants (though not reporting data in support of this statement).
152 However many experiments have since confirmed his conjecture, with the effect verified in adult
153 participants, as well as infants and young children (Maurer et al., 2006; Ozturk et al., 2013). A meta-
154 analysis by Fort et al. (2018) found a modest but consistent *bouba-kiki* effect across a wide range
155 of experimental paradigms and languages, with greater sensitivity for '*bouba*-type' (round shape)
156 than '*kiki*-type' (spiky shape) sound symbolism, a finding that is seen in adults too (Jones et al.,
157 2014). Sensitivity to the kiki effect emerges with age (Tzeng et al., 2017), suggesting that sound
158 symbolism results from both innate and learned mechanisms; a point we will return to below.
159 There are however studies in which the *bouba-kiki* effect has failed (Styles and Gawne,
160 2017; Rogers & Ross, 1975), with likely more unpublished datasets also not showing the effect.
161 Styles and Gawne (2017) argue that failure occurs when the test words used do not conform to
162 the sound structure of the target language, suggesting that the legitimacy of word form may
163 influence the effect. It is also questionable whether the results of forced-choice tasks in
164 experimental settings can be extrapolated to natural language, as the context of the experiment
165 itself may increase participants' awareness of shared properties between stimuli that they would
166 not normally have considered (Sidhu & Pexman, 2018). In addition, Cuskley, Simner & Kirby
167 (2017) argue that the *bouba-kiki* effect is not purely phonological but also significantly influenced
168 by orthography (the shape of symbols used in written language), since most study participants are
169 literate, and the words '*bouba*' and '*kiki*' visually resemble their respective shapes. Similarly, Koriat
170 & Levy (1977) proposed that orthographic systems themselves embody sound-shape symbolism,
171 with letters associated with 'round' sounds (e.g. bilabial sounds or rounded vowels) often appearing
172 more curved for example. However, there is also robust evidence that the effect is independent

from orthography. For example, Bremner et al. (2013) noted the effect in the non-literate Himba population of Northern Namibia (albeit at a reduced incidence of 82%). The most comprehensive study to date however is by Ćwiek et al., (2022), who analysed the orthographic shape of the words 'bouba' and 'kiki' in 25 languages representing 9 language families and 10 writing systems, and found that the effect was no stronger for scripts utilising more rounded forms for '*bouba'* and spikey forms for '*kiki'*. Therefore, while orthography is arguably likely to impact sound-shape symbolism to some degree, particularly in literate participants using written symbols, there is strong evidence that shape-symbolism arises independently of orthographical associations between sound and shape. The existing evidence also strongly supports the long-held hypothesis that the *bouba/kiki* effect is robust cross-culturally (see Davis, 1961).

Yet, some cross-cultural differences were identified by Chen et al., (2016). These authors applied a novel approach of manipulating radial frequency (RF) patterns to investigate perceptual differences between North American and Taiwanese participants. An increase in frequency, amplitude, or spikiness of the sinusoidal modulations of the RF pattern all increased the likelihood of all participants matching the pattern to '*kiki'* instead of '*bouba'*. However, the North American's matching was more dependent upon the amplitude than the Taiwanese, whereas the matching of the Taiwanese was weighted more towards the spikiness of the modulations than the North Americans. This effect could be attributed to the holistic and analytic visual perceptual processes of Eastern and Western culture respectively (Nisbett & Miyamoto, 2005), demonstrating that although the phenomenon may be universal (Ramachandran & Hubbard, 2001) there may be context-induced differences in sensitivity to sound symbolism.

There is still no consensus on which phonemes are most responsible for generating the matching biases, with some suggesting that vowel content causes the effect (Maurer et al., 2006; Ramachandran & Hubbard, 2001) and others highlighting the importance of consonants (Nielsen & Rendall, 2011; Westbury, 2005). It is thought that the close back rounded vowel /u/ naturally matches the rounded figure, but there are also differences in spectral density and attack of the

consonants in '*bouba'* and '*kiki'* such that the stop-consonant 'k' sounds harsher that the bilabial consonant 'b'. Nielsen & Rendall (2013) found evidence that both consonants and vowels are responsible for a portion of the matching bias, and that sound-shape symbolism more generally involves several categories of phonemes. Consistent with this, Westbury et al. (2018) examined the subjective fit of 8000 randomly generated nonwords to roundness or spikiness and found the phonemes /oʊ/ (as in *soak*), /u/, /b/, /m/ and /ɑ/ to be associated with roundness, and the phonemes /t/, /k/, /z/, /i/, and /ɪ/ (as in *sing*) to correspond to spikiness. Extending this to real words, Sidhu et al. (2021) conducted a large-scale study of shape-sound symbolism in English nouns, and found four phonemes associated with roundness (/u/, /m/, /oʊ/, /b/) that were more common in words referring to round objects, and three phonemes associated with spikiness (/k/, /t/, /l/) that were more common in words referring to spiky objects.

2.3. Other Phoneme-Feature Associations

Although much research on sensory sound symbolism has focused on the *sound-size effect* and *sound-shape effect*, several other phoneme-feature associations have been described in the literature (see Lockwood and Dingemanse (2015) for a detailed review). For instance, Vainio & Vainio (2021) describe the phenomenon of sound-action symbolism, in which vocal sounds are associated with a particular body action. An example is the sound-grip effect, in which a precision grip action is associated with front-close vowels and voiceless stop consonants, and a power grip is associated with low-back vowels. This is a form of sound symbolism in which certain vocal signs have connections to motor, perceptual and conceptual representations of particular hand actions. In another study, Hirata et al., (2011) observed that lightness influences sound sensitivity – participants were more successful at identifying consonants when they experienced congruent sound–light pairings (e.g., voiceless consonants with light visual stimuli and voiced consonants with dark visual stimuli) compared to incongruent pairings. A further example of phoneme-feature association comes from a study which investigated how front and back vowels affect conceptual precision (Maglio et al., 2014). Participants showed greater precision in geographic and action

descriptions for labels containing front vowels compared to back vowels. Back vowels in product names however made participants focus on long-term benefits rather than immediate features. This evidence highlights how different vowel sounds can influence mental representation.

Sound symbolic relationships have also been noted for features such as taste (Gallace et al., 2011, Simner et al., 2010), colour (Johansson et al., 2020), and perception of dominant or submissive body postures (Auracher, 2017), suggesting that sound symbolism in language extends far beyond the takete/maluma effect described by Köhler. There is even evidence to suggest that people may draw upon knowledge of form-meaning resemblances in open-ended situations (Davis et al., 2019). When asked to 'draw a creature' described by nonce words, participants included similar elements in drawings, demonstrating that sound symbolism evokes expected properties of referents. These are only some of the examples of phoneme-feature associatinos described in the literature. Mechanisms of such associations and their consequences, both behavioural and evolutionary, are explored in Sections 3 and 4.

3. Mechanisms of Sound Symbolism

The literature on the mechanisms of sound symbolism largely centres around cross-modal correspondences, defined as 'a compatibility between attributes or dimensions of a stimulus (i.e., an object or event) in different sensory modalities' (Spence, 2011, pg.3). The mechanisms of sound symbolism have been comprehensively reviewed by Sidhu & Pexman (2018), who evaluated five proposals, along with their supporting evidence. We summarise these briefly below and discuss additional evidence for each mechanism where appropriate.

According to Sidhu & Pexman, (2018) one way that sound-symbolic association may be explained is through *statistical regularities and co-occurrences* between phonetic features and associated stimuli in the environment. This may be understood via Bayesian integration theory (Spence, 2011), whereby humans combine prior knowledge and sensory information to make cross-modal associations about stimuli. For example, pointed objects tend to produce less tonal sounds with more abrupt transitions whereas rounded forms produce more continuous, tonal sounds, and

these statistical co-occurrences could be reflected in language (Mccormick et al., 2015). Recent computational modelling by Fort & Schwartz (2022) indeed implies that the *bouba/kiki* effect might rely on acoustic cues of spectral balance and temporal continuity. They suggest that cognitively, a sound may be perceived as either spikey or round because it is likely to be produced by a spikey or round object hitting/rolling on a hard surface. Rounded objects, when compared with spiky objects, have lower frequency acoustic resonance modes, meaning that they produce sounds with more energy in the low-frequency part of the acoustic spectrum. They also have smoother trajectories when rolling on a hard surface and therefore produce more continuous acoustic envelopes. These properties, it is argued, may contribute to the formation of cross-modal correspondences responsible for human sensitivity to sound symbolism.

As noted by Sidhu & Pexman (2018), the statistical co-occurrences explanation relies upon experience, and therefore assumes that there is no innate human sensitivity to sound symbolism. Therefore, according to this explanation alone, it might be expected that sound-symbolic effects are not present at birth or in congenitally deaf individuals, yet sound symbolism effects have been observed in 4 month old infants (Ozturk et al., 2013; Peña et al., 2011) and although the effect is diminished in individuals with prelingual auditory deprivation, it is still present at above-chance levels (Gold & Segal, 2020). It is quite possible that statistical learning begins from birth, however it seems plausible that humans may also have an innate sensitivity to sound symbolism. A related view therefore argues that statistical co-occurrences unrerpin sound symbolism, but that these associations are *innate* and not acquired (Ohala, 1994). According to this account, the association between high frequency and smallness, and low frequency and largeness, is innate and universal across species. It is however difficult to generate testable hypotheses to test this proposal, and a distinction is yet to be made between potential innate statistical associations and cross-modal mappings acquired from birth.

Another potential mechanism that also relies on statistical co-occurrences - though this time occurring within language itself - is what Sidhu & Pexman (2018) label *language patterns*. Here,

sound symbolic associations are argued to emerge from co-occurrences between phonological and semantic features, with repeated exposure to phoneme clusters with similar meanings (e.g, phonesthemes) leading to formation of iconic associations. Such patterns can then be used to generate original words, for instance participants using '*gl-'* (as in '*gleam'*, '*glisten'*, '*glow')* to create a novel word related to brightness (Magnus, 2001). However, as noted in section 1.1, it is unclear whether such language patterns can be considered to underpin or emerge from sound symbolic associations.

Sound-symbolic associations have also been explained via the broad mechanism of *shared properties between phonemes and associated stimuli*, which Sidhu & Pexman (2018) argue includes both percerpual links between phonemes and articulatory gestures, as well as higher level conceptual links shared across modalities, e.g. higher-pitch having a connotative association with sharpness. As an example of the former, Sapir (1929) hypothesised that participants might associate high front vowels (e.g., /i/ in 'mil') with small shapes, as the articulation of such vowels requires a smaller oral cavity. In this way, the shape and phoneme share the property of smallness. Shared properties may also be used to explain the bouba/kiki effect, as the pointed lines in the '*kiki'* shape mimic sharp phonemic inflections of the sound '*kiki*', as well as sharp inflection of the tongue on the palate (Ramachandran & Hubbard, 2001). Similarly, the round shape may be labelled '*bouba'* due to the corresponding roundness of mouth and lips when producing the phoneme /u/. In this way, there is a physical relationship between mouth shape and referent object shape, which may provide grounds for resemblance-based association. Ramachandran & Hubbard (2001) suggested that this may occur via cross-modal cortical connections between representations of mouth shapes in motor areas and phonemic representations in proximal auditory regions of the brain. These sensorimotor mappings are proposed to be similar to those in synaesthesia (where a stimulus in one modality triggers the expected sensory experience, as well as activation in at least one other modality) and facilitate non-arbitrary links between an object's visual form and its neural representation. The authors argued that the angular gyrus is important for this cross-modal

association, as it is well-positioned between temporal, parietal, and occipital lobes. They also provide neuropsychological evidence of a patient with left angular gyrus damage that did not exhibit the bouba/kiki effect. Ramachandran and Hubbard's proposal was questioned by Ikegami & Zlatev (2007), who argued that since shape symbolism is not involuntary and often not consciously perceived, it is fundamentally distinct from synaesthesia. The shared properties mechanism is also explored in the Front Oral Cavity (FOC) theory (Masuda, 2007), which proposes that the front oral cavity resonance frequency (which is high for high front vowels) may be the acoustic basis for an association between sound and meaning, and that kinaesthetic factors, such as tongue movement, may contribute to the effect. The FOC theory encompasses acoustic and articulatory bases, providing a possible explanation as to why deaf subjects are still sensitive to sound symbolism (Eberhardt, 1940) – they may feel the size of the oral cavity, and use other visual factors such as the degree of lip rounding.

4. Sound Symbolism in Language Acquisition and Language Evolution

Having provided an overview of the key forms of sound symbolism, and the proposals of their mechanisms discussed in the literature, we now turn to exploring the possible roles of sound symbolism and iconicity in both language acquisition and in the context of language evolution. We also use this evidence to introduce a novel framework to explain how iconicity might fit into the neurobiological framework for the human language processing.

4.1. Sound Symbolism in Language Acquisition

The advantages of sound symbolism in language acquistion have been well documented. It has been found that words learned earlier tend to be more iconic, suggesting that sound symbolism may support word learning (Perry et al., 2015). A study by Imai et al., (2008) suggested that iconicity facilitates early verb learning, and experiments in adults have shown that sound symbolism in natural language may aid word learning too (Lockwood et al., 2016). Sound symbolism may facilitate word learning via

329 'grounding', whereby resemblance based form-meaning relationships allow for shared
330 understanding and the establishment of communication (Cuskley & Sommer, forthcoming). For
331 example, if iconic mappings are shared (such as e.g., /i/ and smallness), when combined with other
332 multimodal inputs (e.g. pointing) this may facilitate shared attention and generation of a form-
333 meaning relationship. A shared understanding of intuitive form-meaning associations may thus
334 facilitate the formation of a lexicon – in other words sound symbolism may act as a scaffold for
335 mapping semantic information, thus bootstrapping word learning (Imai & Kita, 2014; Monaghan
336 et al., 2014). Thus, sound symbolism may aid bootstrapping via establishing 'referentiality' (an
337 ability to map linguistic form to meaning) – a process fundamental to language learning, in
338 combination with Hebbian learning and joint attentional processes (Perniss & Vigliocco, 2014). In
339 a recent review, Nielsen & Dingemanse (2021) found strong evidence for the role of sound
340 symbolism in 'local learning enhancement' - where resemblance-based associations of certain
341 lexical items influence the learning of those items - but not for 'general learning enhancement',
342 where they influence the later learning of arbitrary items. In addition, it has been suggested that
343 iconicity aids the comprehension of communicative signs (Perniss & Vigliocco, 2014) and that the
344 imitative, performative nature of iconic words makes communication more vivid (Lockwood &
345 Dingemanse, 2015). It has however also been noted that more research is required on these effects
346 of sound symbolism, as there may be cross-linguistic variation in learning benefits.

347 **4.2. Iconicity in Language Evolution**

348 Whilst the role of sound symbolism in language acquisition has been well established, its
349 role in the evolution of language is a matter of continuing debate. The beginning of use of words
350 as a communicative means was arguably a pivotal point in human evolution, with small
351 articulations permitting rapid and efficient encoding of information, as well as wide and distant
352 broadcast and communication without sight. However, given the relatively short period of
353 recorded human history, a cross-species comparison is necessary to answer questions relating to
354 language evolution, and the role of iconicity in this context.

355 The core platform for language is face-to-face communication, as this is how languages
356 are learnt and most used. Manual gesture is seen as a likely evolutionary precursor to vocal
357 communication (Rizzolatti & Arbib, 1998), and iconicity has been considered as a logical entry
358 into the language system (Armstrong, 1983), perhaps acting as a bridge between manual (gesture)
359 communicative systems and the verbal language we use today. For example, Levinson & Holler
360 (2014) propose an evolutionary stratification of human communicative systems, whereby the
361 different layers (e.g. joint attention, iconic gesture, turn-taking, and vocalisations) vary in antiquity.
362 Declarative pointing is a form of signalling which facilitates mutual gaze to objects and thus allows
363 for joint attention, while iconic gestures allow the effective depiction of motion, size and spatial
364 relations between objects, such as an offering of something presented with the hand or other body
365 parts (Liebal et al., 2006) or a request made with an open hand (Pollick & De Waal, 2007). The
366 repeated use of iconic gestures would have arguably facilitated the grounding and memorisation
367 of representations, leading to conventionalisation and hence increasing levels of abstraction
368 (Garrod et al., 2007). In this way, iconicity may have been important for achieving *displacement*, in
369 other words the ability to refer to things that are spatially and/or temporally remote (Perniss &
370 Vigliocco, 2014). Displacement would be required to allow hominins to progress from a
371 communication system based on functional reference and pointing to a system based on
372 conceptual reference, and iconicity may therefore have contributed to the development of the
373 cognitive ability required to use conceptually referential signals.
374 While the hypothesis above argues for gesture as a precursor to language in the vocal
375 domain (e.g. Rizzolatti & Arbib, 1998), Perniss et al. (2010) reject this claim, proposing instead
376 that language in manual and vocal modalities must have co-evolved, as linguistic and imagistic
377 components are tightly integrated. This embodied link between linguistic form and sensorimotor
378 experience is argued to reduce cognitive ability needed to unite signs and referents, with linguistic
379 form activating the same systems used in perception and action. The finding that ideophones are
380 more common in narrative contexts and occur alongside iconic gesture (Dingemanse, 2013) is

taken as evidence for embodiment in language, with iconicity and gesture contributing to a multimodal act of depiction. According to this account, the innateness of embodiment is shown by close connections between the hand and mouth in the somatotopic organisation of the human motor cortex (Meier et al., 2008) and in congenitally blind individuals who gesture while speaking despite never having observed this (Iverson & Goldin-Meadow, 2001).

There is also increasing evidence that non-human primates may utilise the processing advantage conveyed by embodiment in the form of iconic gestural communication (Byrne et al., 2017). This evidence comes from studies in which orangutans and chimpanzees have shown the ability to use iconic gesture (pantomime) to depict objects and mime actions related to their use, often elaborating on gestures that fail to elicit the desired response from the recipient. This has been seen in both great apes raised in captivity (e.g. Tanner et al., 2006; Miles et al., 1996), as well as forest-living rehabilitant orangutans (Russon & Andrews, 2011). However, the evidence for the use of iconic gestures in apes is not unequivocal, and other authors have argued that non-human primates are not able to gesture iconically (see Tomasello & Call, 2019). Adding nuance to this debate, Perlman et al., (2012) point to a continuity between gesture and instrumental action in apes, suggesting that gestures are influenced by immediate physical and social contexts and can be adapted into spontaneous iconic gestures on-the-spot.

Of course**,** the great difficulty when examining the relationship between form and meaning from an evolutionary perspective is that it is highly contested whether animal signals can be said to have meaning (i.e. whether non-humans cognitively represent referents; Moore, 2014; Rendall, Owren & Ryan, 2009; Scott-Phillips, 2015). This has led some authors to develop alternative cross-species approaches to examining arbitrariness and iconicity, such as Watson et al.'s (2022) 'optionality' framework, which highlights the presence of linguistic features of arbitrariness in non-human communication. Others (Fischer & Price, 2017) argue that non-human primates do not express communicative or informative intent, and so non-human communication (vocalisations and manual gesture) should be conceived as goal-directed behaviour only, with nothing more than

a probabilistic causal link between the signifier and signified. For example, a chimpanzee may wave an arm in a 'beckoning' gesture to another simply because this produces the desired response in the recipient. Warren & Call (2022) however argue that non-human primates may be capable of applying social inferences to a communicative act, and that there may be mentalistic processes which underlie outcomes. Their model of 'inferential communication' argues that visual perspective taking and knowledge attribution are required for communicative exchange, helping to bridge the gap between animal and human communication.

Yet another useful approach to analyzing the role of iconicity in language evolution is to consider behavioural imitation of others. The human ability to imitate the actions of others is fundamental to turn-taking, which is one of the critical foundations of language evolution (Levinson & Holler, 2014). Imitation is supported by mirror neurons, which code for manual goal-directed movement and fire both when an individual performs a manual task as well as when watching another individual perform the same task. They allow recognition of another's action, as the same neural activation necessary to produce an action is generated via observation. This is important for the development of mutual understanding and an ability to share meaning. Iconicity and the mirror neuron system may therefore contribute to the emergence of expressions, as is seen in signed languages where new signs emerge from iconic gesture (Ahlner & Zlatev, 2010). In non-human primates, mirror neurons have been found in the ventral premotor cortex – comparable to the location of the mirror neuron system in humans - and are argued to have provided a bridge between iconic gesture and the imitation of behaviour (Rizzolatti & Arbib, 1998; see also Ramachandran & Hubbard, 2001). There is also evidence that chimpanzees and other great apes show imitative abilities beyond those of other animals (Bates & Byrne, 2010), which challenges the long-held view that humans are 'imitators' but non-human primates are 'emulators' (e.g. Tomasello et al., 1987; Tomasello, 1996; Tennie et al., 2006). It is argued that similarities in cultural transmission across species indicate our common ancestor was capable of imitating actions with

sufficient fidelity to transmit culturally variant behaviours within and between communities (Whiten et al., 2009).

Pulling all this evidence together therefore suggests that speech may have evolved from proto-dialogue between individuals based on iconic hand gestures and imitation, supported by mutual action recognition facilitated by the mirror-neuron system (Rizzolatti & Arbib, 1998). While sound-symbolism in contemporary language may be perceived as a vestige of the iconic protolanguage (Kita et al. 2010), its manifold roles and advantages in the context of language evolution arguably imply a more integral role within the human language system.

4.3. Iconicity Within the Neurobiological Framework for Human Language

According to one promiment model, the Dual Neurobiological Systems Hypothesis (DNS, Marslen-Wilson & Tyler, 2011), modern human communicative capabilities are controlled by joint activation of bihemispheric (BH) and left-lateralised neural systems, which interact but are functionally and evolutionarily distinguishable (Marslen-Wilson & Bozic, 2018). Broadly speaking, the left-lateralised system encompasses a network of left-hemisphere frontal and temporal regions surrounding the Sylvian fissure and is unique to humans and responsible for supporting complex syntactic functions. The bihemispheric system invloves a broad network of fronto-temporal regions in both hemispheres. It underpins social communication, through the processing of sound-to-meaning mapping, pragmatics, linear adjacency and multimodal interpretation (Bozic et al., 2010; 2015; Marslen-Wilson & Bozic, 2018) – with its capacity for interpretation of multimodal social communication cues particularly relevant for the current context. The bihemispheric system is argued to be evolutionarily primary, with the evidence showing that this system and its functions are highly conserved in non-human primates (Ghazanfar et al., 2008; Seyfarth & Cheney, 2017; Wilson et al., 2015). As such, the bihemispheric system can be considered a promising analogue to the one present in humans at the early stages of language evolution. Given that most forms of iconicity are underpinned by 'linguistic cross-modal correspondence' (Cuskley & Sommer, forthcoming) - excluding only the most direct form-meaning associations such as onomatopoeia

that remain within one sensory modality- and the presence of cross-modal associations has also been noted in non-human primates (Ludwig et al., 2011), it is possible that iconicity may have played a role in bridging the gap between the gestural or imitative visual signs used in early communication and the lexicalised concepts that emerged later. According to this view, cross-modal transfer may have developed into a more sophisticated cross-modal cognitive suite that provided humans with the neurological and behavioural architecture capable of storing and increase cross-modal representations, thus affording us the ability to learn arbitrary symbols necessary for the modern linguistic capacity (Cuskley & Sommer, forthcoming). Therefore, iconicity may in part explain the evolutionary gap between the primate and the modern bihemispheric systems, ultimately setting a precedent for advancing communicative abilities that led to the evolution of the human linguistic capability – with the bihemispheric system likely heavily involved in the processing of iconic sounds, in line with evidence for activation in the bilateral superior temporal sulci (STS) in response to sound symbolic words (Kanero et al., 2014). Taking this evidence into account, we argue that placing iconicity within the neurobiological framework of the DNS hypothesis, and the bihemispheric system in particular, could help strengthen the argument that iconicity played a prominent role in early human language evolution.

5. Factors Limiting the Prevalence of Iconicity

If iconicity provides so many advantages, and is believed to be a stepping stone in the evolution of language, why is language still predominantly arbitrary? From an evolutionary perspective, features of language that improve processing and learnability should survive and become more common. Indeed, Jesperson (1922) found evidence of /i/ replacing other vowels in the historical development of words with meanings related to small size. This may have occurred as sound symbolic associations of /i/ made words containing this phoneme and describing small objects more favourable, and more likely to survive in language evolution – leading to the claim that languages should become 'richer and richer in symbolic words'.

There are however factors that limit the type of meanings that can be expressed iconically. Dingemanse et al (2015) and Sidhu (2019) review several such factors. Firstly, iconic words must have a fairly unique meaning, as similar meanings beg similar forms, and so iconicity may lead to ambiguity unless the referent is very distinct. A wholly iconic language would be dominated by words with similar forms and meanings, leading to uncertainties and deficiencies in processing and learnability. Secondly, the amount of sensory information contained within a referent may limit whether or not it can be described via an iconic relationship. Iconic mappings require sensory features, which restricts their use in language systems. Lupyan & Winter (2019) expand on this further, arguing that iconcity is limited in the expression of abstract concepts in particular, as iconic expressions are too strongly linked to sensory features or certain contexts. For example, certain metaphorical extensions found in English and Hebrew are not possible in Israeli Sign Language (ISL) and American Sign Language as the signs are too iconic (Meir, 2010). One example given is that while English allows for the verb "eat" to extend metaphorically (e.g., "The acid ate the iron key"), ISL's sign for "eat" is too specific, depicting a human eating action at the mouth, which restricts such extensions. Meir discusses how this limitation is also seen in spoken languages where iconic sound-related words are less likely to be used metaphorically across sensory dimensions in 'synaesthetic metaphors'. These examples demonstrate how iconicity may limit the meaning of an expression to a particular context, and thus its potential to become more abstract.

Iconicity in language may also be influenced by linguistic laws such as Zipf's law of abbreviation, which predicts a negative relationship between word length and frequency of use (i.e. words used more frequently tend to be shorter and vice versa). Zipf's law, alongside Menzerath's law (which states that longer communicative constructs are composed of shorter parts) captures the information theoretic principle of minimising code length (compression). Both laws are ubiquitous in communicative systems, having been observed in both manual gesture and vocal communications of non-human primates as well as humans (Heesen et al., 2019; Huang et al., 2020). These laws may decrease iconicity in language - as frequency of use increases, the need for

efficiency prevails over iconicity, resulting in 'lexical elaboration' with increasing levels of arbitrariness (Haiman, 1985).

Iconic words, such as ideophones, also seem to resist deep integration into the grammatical apparatus of a language. According to Dingemanse (2017), this reflects conflict between grammatical integration and the strength of iconic form-meaning mappings – if fully integrated, iconic expressions are subjected to processes of reduction and regularisation so iconicity is reduced. This syntactic conventionalisation diminishes the role of iconicity and could perhaps be further evidence for the processing of iconic signals within the bihemispheric system, as it is the specialised left-hemisphere frontotemporal system that provides the principal neurobiological substrate of core, distinctly human-specific, grammatical capacity, (Bozic et al., 2010, 2015) and iconicity seems to resist grammatical integration.

The prevalence of iconicity in language is also limited by the advantages of arbitrariness. According to (Hockett, 1963), the flexibility of signalling afforded by arbitrariness is a fundamental property of language. Arbitrariness allows referral to any possible concept, and unlike iconicity it can allow for easy discrimination between similar entries in a lexicon. This allows for larger lexica to develop (Gasser, 2004) and allows communication about concepts for which direct perceptual grounding is not possible. A highly iconic lexicon with similar phonological forms for similar meanings would lead to high confusability, and decreased communicative effectiveness (Perniss et al., 2010).

6. Conclusions

In summary, the ancient debate over whether the linguistic sign is arbitrary has been clouded by unnecessary polarisation. A resemblance between form and meaning cannot be sufficient for understanding meaning, and conventionalisation of language throughout evolution is evitable and necessary to maintain the efficiency and versatility of language. Hence while much of the historic debate until recent years has been binary, words cannot be entirely arbitrary or

iconic – they fall on a spectrum instead. Iconicity and arbitrariness both convey their own unique advantages in a linguistic system, representing the adaptation of languages to the constraints of needing to link linguistic form with human experience whilst ensuring an effective signal. It is these constraints that drive the evolution and development of linguistic systems, adding further weight to the hypothesis that theories of language (both phylogenetic and ontogenetic) must value iconicity in addition to arbitrariness, and sound symbolism should be regarded as a core principle of language, and not merely a peripheral phenomenon.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298–348. https://doi.org/10.12697/SSS.2010.38.1-4.11

Armstrong, D. F. (1983). Iconicity, Arbitrariness, and Duality of Patterning in Signed and Spoken Language: Perspectives on Language Evolution. *Sign Language Studies*.

Auracher, J. (2017). Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLOS ONE*, *12*(11), e0187196. https://doi.org/10.1371/JOURNAL.PONE.0187196

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438–481. https://doi.org/10.1037/a0023851

Bates, L. A., & Byrne, R. W. (2010). Imitation: what animal imitation tells us about animal cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 685–695. https://doi.org/10.1002/wcs.77

562 Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311.
563 https://doi.org/10.1353/LAN.2004.0056
564 Bozic, M., Fonteneau, E., Su, L., & Marslen-Wilson, W. D. (2015). Grammatical analysis as a
565 distributed neurobiological function. *Human Brain Mapping*, *36*(3), 1190.
566 https://doi.org/10.1002/HBM.22696
567 Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bihemispheric
568 foundations for human speech comprehension. *Proceedings of the National Academy of Sciences of*
569 *the United States of America*, *107*(40), 17439–17444.
570 https://doi.org/10.1073/PNAS.1000531107/SUPPL_FILE/SAPP.PDF
571 Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013).
572 "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but
573 different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172.
574 https://doi.org/10.1016/J.COGNITION.2012.09.007
575 Byrne, R. W., Cartmill, E., Genty, E., et al. (2017). Great ape gestures: Intentional
576 communication with a rich set of innate signals. Animal Cognition, 20(5), 755–769.
577 https://doi.org/10.1007/s10071-017-1096-4
578 Chen, Y. C., Huang, P. C., Woods, A., & Spence, C. (2016). When "bouba" equals "kiki":
579 Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific*
580 *Reports*, *6*. https://doi.org/10.1038/srep26681
581 Cuskley, C., & Kirby, S. (2013). Synaesthesia, cross-modality and language evolution. In J.
582 Simner, & E. M. Hubbard (Eds.), Oxford Handbook of Synesthesia (pp. 869-907). Oxford
583 University Press. https://doi.org/10.1093/oxfordhb/9780199603329.013.0043
584 Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the
585 bouba-kiki effect. *Psychological research*, *81*(1), 119–130. https://doi.org/10.1007/s00426-015-
586 0709-2

Cuskley, C., & Sommer, K. (forthcoming). The Evolution of Linguistic Iconicity and the Cross-Modal Cognitive Suite. *Oxford Handbook of Iconicity in Language*.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., … Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841). https://doi.org/10.1098/RSTB.2020.0390

Davis, C. P., Morrow, H. M., & Lupyan, G. (2019). What Does a Horgous Look Like? Nonsense Words Elicit Meaningful Drawings. *Cognitive Science*, *43*(10), e12791. https://doi.org/10.1111/COGS.12791

Davis, R. (1961). The Fitness Of Names To Drawings. A Cross-Cultural Study In Tanganyika. *British Journal of Psychology*, *52*(3), 259–268. https://doi.org/10.1111/J.2044-8295.1961.TB00788.X

Dingemanse, M. (2013). Ideophones and gesture in everyday speech. *Gesture*, *13*(2), 143–165. https://doi.org/10.1075/gest.13.2.02din

Dingemanse, M. (2017). Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF - Language Typology and Universals*, *70*(2), 363–384.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/J.TICS.2015.07.013

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). *What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages*.

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. *Language and Speech*, *57*(3), 367–393. https://doi.org/10.1177/0023830913507694/ASSET/IMAGES/LARGE/10.1177_0023830913507694-FIG2.JPEG

Eberhardt, M. (1940). I. A summary of some preliminary investigations of the deaf. *Psychological Monographs*, *52*(1), 1–5. https://doi.org/10.1037/H0093460

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 7–7. https://doi.org/10.1167/7.5.7

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/J.NEUBIOREV.2016.10.014

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, *21*(5). https://doi.org/10.1111/DESC.12659

Fort, M., & Schwartz, J.-L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-23623-w

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of "Bouba" and "Kiki": An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience*, *2*(1), 34–46. https://doi.org/10.1080/17588928.2010.516820

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987. https://doi.org/10.1080/03640210701703659

Gasser, M. (2004). The Origins of Arbitrariness in Language A Language Design Task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26).

Ghazanfar, A. A., Chandrasekaran, C., & Logothetis, N. K. (2008). Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *Journal of Neuroscience*, *28*(17), 4457–4469. https://doi.org/10.1523/JNEUROSCI.0541-08.2008

638 Gold, R., & Segal, O. (2020). The Bouba–Kiki Effect in Persons with Prelingual Auditory
639 Deprivation. *Language Learning and Development*, *16*(1), 49–60.
640 https://doi.org/10.1080/15475441.2019.1685386/SUPPL_FILE/HLLD_A_1685386_SM
641 8566.DOCX
642 Haiman, J. (1985). Natural Syntax. Iconicity and Erosion. In *Cambridge Studies in Linguistics* (Vol.
643 44). Cambridge University Press.
644 http://www.jstor.org/action/showPublisher?publisherCode=springer
645 Heesen, R., Hobaiter, C., Ferrer-I-Cancho, R., & Semple, S. (2019). Linguistic laws in
646 chimpanzee gestural communication. *Proceedings of the Royal Society B*, *286*(1896).
647 https://doi.org/10.1098/RSPB.2018.2900
648 Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and
649 visual lightness using Garner's speeded classification task. *Perceptual and Motor Skills, 113*(3),
650 929–940. https://doi.org/10.2466/15.21.28.PMS.113.6.929-940
651 Huang, M., Ma, H., Ma, C., Garber, P. A., & Fan, P. (2020). Male gibbon loud morning calls
652 conform to Zipf's law of brevity and Menzerath's law: insights into the origin of human
653 language. *Animal Behaviour*, *160*, 145–155.
654 https://doi.org/10.1016/J.ANBEHAV.2019.11.017
655 Ikegami, T., & Zlatev, J. (2007). From Non-Representational Cognition To Language. *Body,*
656 *Language and Mind*, *1: Embodiment*, 241–283.
657 Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language
658 acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological*
659 *Sciences*, *369*(1651). https://doi.org/10.1098/RSTB.2013.0298
660 Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb
661 learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: gesture in blind speakers and listeners. *Developmental Science*, *4*(4), 416–422. https://doi.org/10.1111/1467-7687.00183

Jesperson, O. (1922). Symbolic value of the vowel i. *Phil J Comp Phil*, *1*, 15–33.

Johansson, N., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition*, *12*(1), 56–83. https://doi.org/10.1017/LANGCOG.2019.35

Jones, M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J., & Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1119-1144). Cognitive Science Society.

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*(5). https://doi.org/10.1371/JOURNAL.PONE.0097905

Kita, S., Kantartzis, K., & Imai, M. (2010). *Children Learn Sound Symbolic Words Better: Evolutionary Vestige Of Sound Symbolic Protolanguage*. 206–213. https://doi.org/10.1142/9789814295222_0027

Köhler, W. (1929). *Gestalt Psychology*. Liveright.

Koriat, A. & Levy, I. (1977). The symbolic implications of vowels and of their orthographic representations in two natural languages. *Journal of Psycholinguistic Research, 6*(2), 93–104.

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

Liebal, K., Pika, S., & Tomasello, M. (2006). *Gestural communication of orangutans (Pongo pygmaeus)*.

Locke, J. (1690). *An Essay Concerning Human Understanding* (WLC, 2009).

686 Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral,
687 developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*,
688 1246. https://doi.org/10.3389/FPSYG.2015.01246/BIBTEX
689 Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word
690 learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1274–1281.
691 https://doi.org/10.1037/XLM0000235
692 Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high
693 luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans.
694 *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20661–
695 20665. https://doi.org/10.1073/PNAS.1112605108/SUPPL_FILE/SM01.MOV
696 Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't
697 languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences,*
698 *373*(1752), 20170137. https://doi.org/10.1098/rstb.2017.0137
699 Maglio, S. J., Rabaglia, C. D., Feder, M. A., Krehm, M., & Trope, Y. (2014). Vowel sounds in
700 words affect mental construal and shift preferences for targets. *Journal of Experimental*
701 *Psychology: General, 143*(3), 1082–1096. https://doi.org/10.1037/a0035543
702 Magnus, Margaret. (2001). *What's in a word?: Studies in phonosemantics*. 202.
703 https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/243679
704 Marslen-Wilson, W. D., & Bozic, M. (2018). Dual neurobiological systems underlying language
705 evolution: inferring the ancestral state. *Current Opinion in Behavioral Sciences*, *21*, 176–181.
706 https://doi.org/10.1016/J.COBEHA.2018.05.004
707 Marslen-Wilson, W. D., & Tyler, L. K. (2011). The Communicative Brain. *Res Horiz*, *16*, 8–9.
708 Masuda, K. (2007). *The physical basis for phonological iconicity* (pp. 57–71).
709 https://doi.org/10.1075/ill.5.08mas

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

Mccormick, K., Kim, J. Y., List, S., & Nygaard, L. C. (2015). Sound to meaning mappings in the Bouba-Kiki effect. *CogSci*, Vol. 2015, pp. 1565-1570. https://www.researchgate.net/publication/309485703

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press.

Meier, J. D., Aflalo, T. N., Kastner, S., & Graziano, M. S. A. (2008). Complex Organization of Human Primary Motor Cortex: A High-Resolution fMRI Study. *Journal of Neurophysiology*, *100*(4), 1800. https://doi.org/10.1152/JN.90531.2008

Meir, I. (2010). Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language, 86*(4), 865–896. https://doi.org/10.1353/lan.2010.0044

Miles, H. L., Mitchell, R. W., & Harper, S. E. (1996). Simon says: The development of imitation in an enculturated orangutan. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 278-299). Cambridge University Press.

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

Moore, R. (2014). Ape Gestures: Interpreting Chimpanzee and Bonobo Minds. *Current Biology*, *24*, R645–R647. https://doi.org/10.1016/j.cub.2014.06.001

Newman, S. S. (1933). Further Experiments in Phonetic Symbolism. *The American Journal of Psychology*, *45*(1), 53. https://doi.org/10.2307/1414186

Nielsen, A. K. S., & Dingemanse, M. (2021). Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339/ASSET/IMAGES/LARGE/10.1177_0023830920914339-FIG1.JPEG

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*(2), 153–163. https://doi.org/10.1037/a0030553

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*(2), 115–124. https://doi.org/10.1037/A0022268

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473. https://doi.org/10.1016/J.TICS.2005.08.004

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound Symbolism*, 325–347.

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

Peirce, C. S. (1974). *Collected papers of Charles Sanders Peirce* (6th Edition). Harvard University Press.

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

Perlman, M., Tanner, J. E., & King, B. J. (2012). A mother gorilla's variable use of touch to guide her infant. In S. Pika & K. Liebal (Eds.), Developments in primate gesture research (pp. 55–72). John Benjamins Publishing. https://doi.org/10.1075/gs.6.04per

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in Signed and Spoken Vocabulary: A Comparison Between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/FPSYG.2018.01433

761 Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a General Property of
762 Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*(DEC).
763 https://doi.org/10.3389/FPSYG.2010.00227
764 Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the
765 experience of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological*
766 *Sciences*, *369*(1651). https://doi.org/10.1098/RSTB.2013.0300
767 Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in english and Spanish and its relation
768 to lexical category and age of acquisition. *PLoS ONE*, *10*(9).
769 https://doi.org/10.1371/journal.pone.0137147
770 Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain*
771 *Sciences*, *13*(4), 707–784. https://doi.org/10.1017/S0140525X00081061
772 Pollick, A. S., & De Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the*
773 *National Academy of Sciences of the United States of America*, *104*(19), 8184–8189.
774 https://doi.org/10.1073/PNAS.0702624104
775 Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia-A Window Into Perception,
776 Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34.
777 Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses
778 and theoretical implications. *Cognitive Science*, *31*(1), 157–168.
779 https://doi.org/10.1080/03640210709336988
780 Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*,
781 *78*(2), 233–240. https://doi.org/10.1016/J.ANBEHAV.2009.06.007
782 Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5),
783 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0
784 Rogers, S. K., & Ross, A. S. (1975). A cross cultural test of the Maluma Takete phenomenon.
785 *Perception*, *4*(1), 105–106. https://doi.org/10.1068/P040105

Russon, A., & Andrews, K. (2011). Orangutan pantomime: elaborating the message. *Biology Letters*, *7*(4), 627–630. https://doi.org/10.1098/RSBL.2010.0564

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225–239. https://doi.org/10.1037/H0070931

Saussure, F. de. (1959). *Course in General Linguistics*.

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, *5*(FEB), 80. https://doi.org/10.3389/fpsyg.2014.00080

Scott, J. H., Foley, R. A., & Bozic, Mirjana. (n.d.). Iconic Labels Enhance Visual Perception - Implications for Language Evolution. *Frontiers In Psychology*.

Scott-Phillips, T. C. (2015). Meaning in animal and human communication. *Animal Cognition*, *18*(3), 801–805. https://doi.org/10.1007/S10071-015-0845-5

Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic Bulletin and Review*, *24*(1), 79–84. https://doi.org/10.3758/s13423-016-1059-9

Sidhu, D. M. (2019). *Explorations of Sound Symbolism and Iconicity* (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca.

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

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin and Review*, *28*(4), 1390–1398. https://doi.org/10.3758/S13423-021-01883-3

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception, 39*(4), 553–569. https://doi.org/10.1068/p6591

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics 2011 73:4*, *73*(4), 971–995. https://doi.org/10.3758/S13414-010-0073-7

812 Styles, S. J., & Gawne, L. (2017). When Does Maluma/Takete Fail? Two Key Failures and a
813 Meta-Analysis Suggest That Phonology and Phonotactics Matter. *I-Perception*, *8*(4).
814 https://doi.org/10.1177/2041669517724807
815 Svantesson, J. O. (2017). Sound symbolism: the role of word sound in meaning. *Wiley*
816 *Interdisciplinary Reviews: Cognitive Science*, *8*(5), e1441. https://doi.org/10.1002/WCS.1441
817 Tanner, J. E., Patterson, F. G., & Byrne, R. W. (2006). The development of spontaneous gestures
818 in zoo-living gorillas and sign-taught gorillas: From action and location to object
819 representation. *Journal of Developmental Processes, 1,* 69-103.
820 Tennie, C., Call, J., & Tomasello, M. (2006). Push or Pull: Imitation vs. Emulation in Great Apes
821 and Human Children. *Ethology*, *112*(12), 1159–1169. https://doi.org/10.1111/J.1439-
822 0310.2006.01269.X
823 Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded
824 function. *Quarterly Journal of Experimental Psychology*, *64*(12), 2392–2404.
825 https://doi.org/10.1080/17470218.2011.605898/ASSET/IMAGES/LARGE/10.1080_17
826 470218.2011.605898-FIG2.JPEG
827 Tomasello, M. (1996). Do apes ape? In C. M. Heyes & B. G. Galef, Jr. (Eds.), *Social learning in*
828 *animals: The roots of culture* (pp. 319–346). Academic Press. https://doi.org/10.1016/B978-
829 012273965-1/50016-9
830 Tomasello, M., & Call, J. (2019). Thirty years of great ape gestures. Animal Cognition, 22(5),
831 461–469. https://doi.org/10.1007/s10071-018-1167-1
832 Tomasello, M., Davis-Dasilva, M., Camak, L., & Bard, K. (1987). Observational learning of tool-
833 use by young chimpanzees. *Human Evolution, 2,* 175-183.
834 https://doi.org/10.1007/BF02436405
835 Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). The Specificity of Sound Symbolic
836 Correspondences in Spoken Language. *Cognitive Science*, *41*(8), 2191–2220.
837 https://doi.org/10.1111/COGS.12474

838 Vainio, L., & Vainio, M. (2021). Sound-Action Symbolism. *Frontiers in Psychology*, *12*, 3848.
839 https://doi.org/10.3389/FPSYG.2021.718700/BIBTEX
840 Warren, E., & Call, J. (2022). Inferential Communication: Bridging the Gap Between Intentional
841 and Ostensive Communication in Non-human Primates. *Frontiers in Psychology*, *12*, 718251.
842 https://doi.org/10.3389/FPSYG.2021.718251
843 Watson, S. K., Filippi, P., Gasparri, L., Falk, N., Tamer, N., Widmer, P., Manser, M., & Glock,
844 H. J. (2022). Optionality in animal communication: a novel framework for examining the
845 evolution of arbitrariness. *Biological Reviews of the Cambridge Philosophical Society*, *97*(6), 2057–
846 2075. https://doi.org/10.1111/BRV.12882
847 Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference
848 task. *Brain and Language*, *93*(1), 10–19. https://doi.org/10.1016/J.BANDL.2004.07.006
849 Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for
850 sound symbolism: Distributional properties predict cue strength. *Journal of Memory and*
851 *Language*, *99*, 122–150. https://doi.org/10.1016/J.JML.2017.09.006
852 Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation,
853 over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of*
854 *the Royal Society B: Biological Sciences*, *364*(1528), 2417.
855 https://doi.org/10.1098/RSTB.2009.0069
856 Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., Thiele, A., Griffiths, T. D.,
857 Marslen-Wilson, W. D., & Petkov, C. I. (2015). Auditory sequence processing reveals
858 evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature*
859 *Communications*, *6*. https://doi.org/10.1038/ncomms9901
860 Zangenehpour, S., & Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex
861 following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, *48*(2), 591–600.
862 https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2009.10.022
863