Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language

Marcus Perlman and Ashley A. Cain
University of Wisconsin-Madison / San José State University

Scholars have often reasoned that vocalizations are extremely limited in their potential for iconic expression, especially in comparison to manual gestures (e.g., Armstrong & Wilcox, 2007; Tomasello, 2008). As evidence for an alternative view, we first review the growing body of research related to iconicity in vocalizations, including experimental work on sound symbolism, cross-linguistic studies documenting iconicity in the grammars and lexicons of languages, and experimental studies that examine iconicity in the production of speech and vocalizations. We then report an experiment in which participants created vocalizations to communicate 60 different meanings, including 30 antonymic pairs. The vocalizations were measured along several acoustic properties, and these properties were compared between antonyms. Participants were highly consistent in the kinds of sounds they produced for the majority of meanings, supporting the hypothesis that vocalization has considerable potential for iconicity. In light of these findings, we present a comparison between vocalization and manual gesture, and examine the detailed ways in which each modality can function in the iconic expression of particular kinds of meanings. We further discuss the role of iconic vocalizations and gesture in the evolution of language since our divergence from the great apes. In conclusion, we suggest that human communication is best understood as an ensemble of kinesis and *vocalization*, not just speech, in which expression in both modalities spans the range from arbitrary to iconic.

Keywords: iconicity, language evolution, modality, sound symbolism, vocalization

The linguist Charles Hockett suggested that spoken languages are forced to be arbitrary because of the vocal channel's intrinsic limitations for iconic expression (Hockett, 1978). As he explained (p. 274):

Gesture 14:3 (2014), **320–350.** DOI 10.1075/gest.14.3.03per
ISSN 1568–1475 / E-ISSN 1569–9773 © John Benjamins Publishing Company

> When a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out. In one-dimensional projection, an elephant is indistinguishable from a woodshed. Speech perforce is largely arbitrary.

Thus Hockett proposes that vocalizations lack the potential to convey semantic information through their form, beyond their arbitrary, learned associations. In the present study we examine this claim, and more specifically, whether people are able to produce iconic vocalizations to convey different kinds of meanings. As we shall see, the vocal modality can be used iconically to a much greater degree than has often been supposed. We suggest that this has consequences for how the role of iconicity in gesture, relative to that of vocalization, is assessed in theories of language origins.

We adopt a comprehensive definition of iconicity, using the term to refer, across modalities and media, to any sort of correspondence between the form of an expression and its meaning (cf. Peirce, 1955; Perniss & Vigliocco, 2014; Wescott, 1971). Iconic forms can vary in the degree of abstraction involved in their representation of meaning, and they may incorporate processes of metonymy and metaphor, including correspondences across modalities. For example, with respect to the task to distinguish an elephant from a woodshed, one could communicate the idea of an elephant by a visible iconic gesture, such as extending one's arm from one's face and moving it up and down to represent the animal's trunk (e.g., https://web.archive.org/web/20151030192220/https://www.youtube.com/watch?v=FPmbZVeOBL0). This video also shows how one can express the idea of an elephant by producing a vocal imitation of the animal's roar. (Indeed, it may sometimes be most natural or effective to combine these two modalities of iconic expression together.)

One can see varying levels of abstraction in the iconicity in these two examples. The imitation of an elephant vocalization is an attempt to produce a veridical representation of the animal's sound. In comparison, the use of one's arm to represent the elephant's trunk is more abstract. The arm is long and thin like a trunk, but clearly differs in finer details of color, texture, and shape. Sound-based iconicity can be abstract too, such as in the musical accompaniment in this cartoon of an elephant and a mouse (https://web.archive.org/web/20151030193927/https://www.youtube.com/watch?). The musical sound effects of the trombone and bassoon reflect qualities of the elephant's lumbering movement through their tempo and pitch (also note the sound effects of the mouse climbing the ladder at 49 s). Yet the representation is abstract, not intended as an accurate imitation of the sound of an elephant's footsteps.

Whereas it has been widely supposed, at least since Saussure (1916), that the forms of words are purely arbitrary in relation to their meanings, speaker gestures and signs in sign language have often been assumed to be iconic. This has suggested to many that language first began as a form of gesturing, since it seems difficult to envisage how the arbitrary pairing of sounds with meaning could have arisen at first (e.g., Arbib, 2012; Armstrong & Wilcox, 2007; Corballis, 2002; Steklis & Harnad, 1976; Tomasello, 2008). According to this theory, iconic gestures are needed to ground the development of symbols — a crucial step in language evolution that cannot be accomplished by inherently arbitrary vocalizations.

This rationale is illustrated in a thought experiment by Tomasello (2008). He asks the reader to imagine two isolated groups of children who grow up together from birth, well cared for somehow, but without any external model of communication. In this hypothetical experiment, one group of children is unable to vocalize and must gesture to communicate, while the other group cannot communicate with their hands and must only use vocalizations. Does each group of children develop a language? And if so, how does that happen?

As Tomasello notes, we know a lot about what happens in the gesture group based on observations of the development of home sign systems and the creation of new signed languages (Armstrong & Wilcox, 2007; Goldin-Meadow, 2003; Sandler, 2013). Gestures are readily used for pointing and directing an interlocutor's attention to something, as well as for pantomiming actions and depicting shapes and spatial layouts. Therefore the children would naturally create deictic and iconic gestures to ground the development of symbolic communication. These gestures would develop first into a rudimentary symbol system, and over a few generations, into a bona fide grammatical signed language. In contrast, Tomasello offers a different assessment of the vocal scenario (2008, p. 228):

> It is difficult to imagine [the children] inventing on their own vocalizations to refer the attention or imagination of others to the world in meaningful ways — beyond perhaps a few vocalizations tied to emotional situations and/or a few instances of vocal mimicry. Humans have no natural tendencies in the vocal modality — analogous to following gaze directionally in space or interpreting actions as intentional in the gestural/visual modality — to serve as starting points. And so the issue of conventionalizing already meaningful communicative acts never arises.

Similar arguments about the limitations of iconicity in the vocal modality are incorporated into a variety of theoretical perspectives, from the embodied (Armstrong & Wilcox, 2007) to the modular (Pinker & Jackendoff, 2005). However, despite many such claims, there is actually considerable evidence suggesting that the vocal modality holds more potential for iconicity than is often realized. Below we review some of this evidence, including experimental work on sound symbolism, cross-linguistic studies documenting iconicity in the grammars and lexicons of languages, and finally some recent experiments that examine iconicity in the production of speech and vocalizations. This research shows that vocal iconicity extends well beyond the limited boundaries of emotional expression and vocal mimicry that are suggested by Tomasello. Moreover, as we discuss further below, the iconic expression of emotion and vocal mimicry is hardly trivial.

Sound symbolism

A multitude of studies on sound symbolism have found that people are often consistent in the meanings they associate with different speech sounds, including various vowels and consonants (for thorough treatments, see Jacobson & Waugh, 1979, Ch. 4; Hinton, Nichols, & Ohala, 1994; Imai & Kita, 2014; Nuckolls, 1999; Wescott, 1971). One of the most studied and commonly cited examples is the association between front and back vowels (e.g., /i/ vs. /u/) and small and large size (Ohala, 1994), as in the English word "teeny" compared to "huge." Standard explanations for this association point to the resonant frequencies of front versus back vowels. Front vowels have a high-pitched second formant compared to the low second formant of back vowels. This correspondence is seen to reflect a natural correlation between pitch and size: big vocal tracts and big things in general tend to produce lower pitched sounds.

A pioneering study in pitch-size symbolism presented participants with spoken pairs of minimally contrasting nonsense words and asked them to indicate which conveyed a larger object (Sapir, 1929). People consistently associated words containing vowels like /a/ with the large object, especially in contrast to the front vowel /i/. In a more recent study, participants named novel objects with nonwords that varied in the number of "large" and "small" phonemes (Thompson & Estes, 2011). The number of large phonemes in the selected name varied according to the size of the object. Experiments show further that consistent sound-meaning judgments extend to a host of meanings that may relate to size, including *gender* and *quickness*, and also attitudes like *affection*, *intimacy*, *disdain*, and *acquiescence* (e.g., Bentley & Varon, 1933; Jespersen, 1933; Newman, 1933; Ohala, 1994).

Another well-examined case of sound symbolism is the so-called "bouba-kiki" effect, which involves the association of particular phonemic patterns with certain qualities of shape (Ramachandran & Hubbard, 2001). Köhler (1929) performed the original experiment in which people decided which of two words "takete" or "baluma" they would choose as the most fitting label for a pointed, angular shape compared to a curvy one. The overwhelming majority of respondents select words like “takete” for the angular shape and “baluma” for the curved shape. More generally, the pattern of voiceless obstruent consonants alternated with front, closed, and unrounded vowels tends to evoke an angular sense, whereas voiced sonorants combined with open, rounded vowels evoke a sense of rounded (Ahlner & Zlatev, 2011). The effect has been replicated with speakers of different languages, including Spanish (Köhler, 1929), Otjiherero (a language spoken by the Himba in Northern Namibia; Bremner et al., 2013), and in 8 to 14-year-old speakers of Swahili and the Bantu language Kitongwe (Davis, 1961).

The kiki-bouba effect has also been demonstrated in more implicit, online tasks. For example, when participants performed a lexical decision task with letter strings containing either stop consonants or continuants presented within the lines of an angular or rounded shape, they were slower to reject non-words when they were presented within a compatible shape (Westbury, 2004). Another experiment found that participants were faster to identify novel pointy or round objects after hearing a sound-symbolic compatible label (Kovic, Plunkett, & Westermann, 2010). Lupyan and Casasanto (2014) also demonstrated that nonce words with phonetic features associated with pointy and rounded shapes facilitated the learning of related categories. Participants were faster and more accurate in learning to categorize two different species of aliens with pointed or rounded heads when they were associated with the respectively compatible labels “crelch” and “foove”.

Experimental studies of sound symbolism extend to other semantic domains too, not just size and shape. For example, Cuskley (2013) asked participants to listen to nonsense CVCV words and adjust the speed of an animated bouncing ball to match the level of speed the words expressed. The stimuli varied according to voicing, vowel quality, and how the CV syllable was reduplicated. The reduplication of consonants with alternating vowels (e.g., “kiku”) led to higher ratings in speed, whereas words with back vowels (e.g., “gugu”) were rated as being particularly slow.

Other studies have examined a much broader array of meanings. A common method, used especially in several early studies, asks participants to match their own language equivalents to the meanings of antonyms from different, unknown languages (Brown et al., 1955; Gebels, 1969; LaPolla, 1994). Participants are often better than chance at identifying matches across numerous meanings.

A few studies have examined people’s semantic associations across large sets of English phonemes with respect to a variety of meanings. Greenberg and Jenkins (1966) asked participants to rate English consonants and vowels along a number of semantic scales, and used factor analysis to place each phoneme within a semantic space. Different categories of consonants were consistently distinguished along several semantic dimensions including *abrupt-continuous*, *liquid-solid*, *tight-loose*, *delicate-rugged*, *angular-rounded*, *active-passive*, *good-bad*, and *inhibited-free*. Vowels, based largely on tongue position, were distinguished along numerous other scales such as *high-low*, *sharp-dull*, *narrow-wide*, *thick-thin*, *oblong-round*, *large-small*, and *falling-rising*.

A marketing study used a similar approach to examine the sound symbolic associations of product brand names, which were constructed to contrast various consonant and vowel characteristics (Klink, 2000; also see Kelly, Leben, & Cohen, 2004). Participants completed a questionnaire with questions like "Which brand of ketchup seems thicker? Nidax or Nodax?" (p. 11). Among a host of findings, names with front vowels were judged as *smaller*, *lighter* (*vs. darker*), *milder*, *thinner*, *softer*, *faster*, *colder*, more *bitter*, more *feminine*, *friendlier*, *weaker*, *lighter* (*vs. heavier*), and *prettier*; fricatives compared to stops were judged as *smaller*, *faster*, *lighter* (*vs. heavier*), and more *feminine*; voiceless stops were *smaller*, *faster*, *lighter* (*vs. heavier*), *sharper*, and more *feminine*; and voiced fricatives were *faster*, *softer*, and more *feminine*.

Research also shows that young children are sensitive to sound symbolism. English speaking children as young as 2.5 years of age are sensitive to the bouba-kiki effect and similar mappings (Maurer, Pathman, & Mondloch, 2006). In a help-the-puppet scenario, children were presented with object labels containing either rounded or unrounded vowels and asked which of two shapes — one rounded and one angular — fit with the label. The children reliably chose the shape-compatible labels.

In addition to shape, Imai and colleagues (2008) showed that Japanese children roughly 2 and 3 years of age could match novel sound symbolic verbs to compatible actions. 3-year-old Japanese children were also better at learning verbs when they were sound symbolic. They succeeded in generalizing the use of novel sound-symbolic verbs for actions, but failed with non-sound-symbolic verbs. Later work with English-speaking 3-year-olds showed that they were also better at generalizing novel sound-symbolic verbs compared to non-sound-symbolic ones (Kantartzis, Imai, & Kita, 2011).

Iconicity in lexicon and grammar

There is also a considerable amount of iconicity evident in the grammars and lexicons of spoken languages.[1] In some cases, such as in certain patterns of grammar, this iconicity can be highly schematic. Greenberg (1966), for instance, observes that the order of elements in language parallels their order in physical experience

[1] One widely acknowledged source is often characterized as *relative* iconicity (Monaghan, Shilcock, Christiansen, & Kirby, 2014; also see Waugh, 2000). However, relative iconicity is internal to a language and does not necessarily involve so-called *absolute* correspondences between form and meaning.

or knowledge. The Latin narrative sequence "*veni, vidi, vici*" is an often-cited example. Another abstract form of iconicity in grammar is the well-established principle that increased morphological complexity tends to correspond to increased semantic complexity. For example, Haiman (1980) notes that positive, comparative, and superlative degrees of adjectives generally show an increase in the number of phonemes.

Many studies also document the prevalence of iconicity in lexicons across the world's languages, which often feature a grammatically distinct class of words termed variously as ideophones, mimetics, and expressives (Diffloth, 1971; Dingemanse, 2012). These words typically serve adverbial or adjectival functions through the depiction of sensory and motor imagery, and can express a wide range of meanings spanning across modalities: animate and inanimate sounds, luminance and color, manner of movement and speed, shape, temporal aspect, emotional and psychological states, size, texture, visual appearance, taste, and temperature. Although languages exhibit this iconic lexical class to different degrees, some scholars propose that it is universal of languages (Diffloth, 1972; Voeltz & Kilian-Hatz, 2001). Large iconic lexicons have been identified in nearly all sub-Saharan African languages (Childs, 1994), in some Australian Aboriginal languages (Alpher, 2001; McGregor, 2001; Schultze-Berndt, 2001), in Japanese, Korean, and Southeast Asian languages (Hamano, 1998; Diffloth, 1972; Watson, 2001), indigenous South American languages (Nuckolls, 1996), and Balto-Finnic languages (Mikone, 2001).

In addition to iconicity in these particular lexical classes, large-scale comparative studies sampling across language families find statistical regularities in the phonological sound shapes that are used to express specific meanings. For example, Ultan (1978) found that languages tend to use high front vowels to express diminutive meanings. Similarly, Tanz (1971) examined the indexical words used to express proximity and distance (e.g., in English, "here" vs. "there"), and found a strong tendency for languages to express close distance with front vowels and far distance with back vowels. Another study showed a preponderance of nasal consonants in words for 'nose' and bilabial consonants in words for 'mouth' (Urban, 2011). There is also great similarity across languages in the form of the interrogative meaning equivalent to "huh?" (Dingemanse, Torreira, & Enfield, 2013). A more comprehensive survey looked for evidence of sound symbolism within 40 basic vocabulary words across 121 language families and 52 isolates and unclassified languages (Wichmann, Holman, & Brown, 2010). Overall these vocabulary items showed a correlation between the form of the words and their meaning, and in particular, meanings like *breast*, *I*, *knee*, *you*, *nose*, *name*, and *we* appear to have similar phonological forms across languages.

Iconicity in vocal production

Thus people are sensitive to sound symbolism, and there is considerable evidence of iconicity fossilized into the grammars and lexicons of languages. However, only a few studies have directly examined iconicity in the dynamic production of speech. Do speakers modulate their speech in ways that reflect iconic mappings between form and meaning?

Some recent research on spoken prosody suggests that speakers do have a natural tendency to modulate their speech in iconic ways. For example, in one set of experiments, participants used prescribed phrases to describe the direction a dot moved on a computer screen — either up, down, left, or right (Shintel, Nusbaum, & Okrent, 2006). Speakers raised their pitch when describing upward movement, but lowered their pitch when describing downward movement. They also increased or decreased their articulation rate when describing the dot as moving fast or slow. Further experiments showed that listeners are sensitive to modulations of speech rate when they interpret the meaning expressed by the utterance (Shintel et al., 2006; Shintel & Nusbaum, 2007), particularly when speed is contextually relevant (Shintel & Nusbaum, 2008).

Speakers also produced iconic modulations in their prosody when reading short stories that contrasted along different elements of meaning. For example, they inflected their pitch when reading stories about *high location* and *upward movement* versus *low location* and *downward movement* (Clark, Perlman, & Johansson Falck, 2014) and about *small* versus *big size* (Perlman, Clark, & Johansson Falck, 2015). They also modulated their articulation rate when reading stories about *fast* versus *slow-paced* events (Perlman et al., 2015), and when providing spontaneous, open-ended descriptions of fast versus slow events viewed in short video clips (Perlman, 2010).

Some research suggests that adults using infant directed speech might be especially inclined to produce modulations in their prosody of an iconic character (Nygaard, Herold, & Namy, 2009). Three speakers were shown a nonce word that represented one of the antonymic meanings *happy*|*sad*, *hot*|*cold*, *big*|*small*, *tall*|*short*, *yummy*|*yucky*, or *strong*|*weak*. Instructed to say, "Can you get the [nonce word] one?" as if to an infant, speakers consistently distinguished antonymic words along particular combinations of acoustic parameters, at the levels of both word and sentence. For example, the meaning *big* was characterized by lower pitch, longer duration, and higher intensity than *small*, while *tall* was characterized by a longer duration and greater pitch variation than *short*.

In a more implicit production experiment (Parise & Pavani, 2011), participants viewed stimuli varying along three dimensions: shape (triangle, hexagon, dodecagon), luminance (white, gray, black), and size (small, medium, large). They performed a go/no-go task in which they produced the vowel /a/ in response to the two outermost exemplars of a given dimension (e.g., black or white but not gray) for as long as the stimulus was on the screen. The results showed that participants modulated the syllable in predictable ways according to the stimuli, for example, articulating the sound with higher intensity when responding to dodecagons compared to triangles and white shapes compared to black ones.

Thus people spontaneously modulate their prosody when speaking, but are they able to generate *novel* iconic vocalizations to express particular meanings? In a few recent studies, participants were tasked to actually create iconic sounds. In one set of experiments, participants played a communication game in which they used manual gestures, vocalizations, or a combination to communicate to a partner words from a shared list that included emotions, actions, and objects (Fay, Arbib, & Garrod, 2013; Fay, Lister, Ellison, & Goldin-Meadow, 2014). Players demonstrated a moderate amount of success in the vocal condition, but they were significantly better with gestures, and gained no added benefit from the combination of modalities. The authors conclude that the iconic, motivated nature of gestures serves to ground the creation of labels, whereas this grounding is not afforded by the more arbitrary nature of vocalizations.

However, a different style of communication game showed more positive results for the iconic potential of vocal communication (Perlman, Dale, & Lupyan, 2014). Pairs of participants played a vocal charades game with nine antonymic pairs of adjectives and adverbs. Over ten rounds, players took turns producing non-linguistic vocalizations to communicate each meaning to their partner. The vocalizations were measured for their average pitch, pitch change, duration, intensity, and harmonics-to-noise ratio. Analysis revealed that players created non-linguistic vocalizations with highly consistent acoustic properties that systematically differentiated between each of the meanings.

Here we report a slightly modified version of this study, greatly expanding the range of meanings to be expressed. In accordance with the results of our previous study, we expect that the players will show consistencies in the acoustic properties of the vocalizations they produce for the different meanings. If this expectation is met this would indicate that players shared intuitions for how the qualities of their voice mapped to each meaning, implicating iconicity in the mappings.

Method

Participants

15 pairs of undergraduate students from the University of California, Santa Cruz participated in the study in exchange for course credit.

Materials

The game was played with 3 by 5 inch, white index cards, each with one of 60 words ("meanings") typed on one side. The meanings consisted of 30 antonymic pairs spanning a mix of basic concepts (see Table 1). None related primarily to sound, but four had spatial meanings that are often extended metaphorically to describe sounds (e.g., *long*|*short*, *up*|*down*).

Design and procedure

Participants played a game similar to the parlor game charades. Pairs of players took turns as the "vocalizer" and the "guesser" as they attempted to communicate the set of target meanings. Vocalizers were only allowed to produce non-linguistic vocalizations to communicate each meaning; no words or bodily gestures were permitted.

When participants arrived, they were greeted by the experimenter, who made introductions and provided instructions. The players were seated about five feet apart, facing each other across a table. Thus they were able to see each other. However, as we shall explain in the Discussion, such facial and bodily expressions that the participants may have unavoidably engaged in, do not affect the observations on acoustic consistency that will be reported. Each participant was given two sets of 15 cards. The cards were randomized except that individuals always had both meanings of an antonymic pair shuffled within their 30 total cards.

On each turn, the vocalizer picked up the next card from the top of the stack, read it, and set it face down. To discourage manual gesturing, the vocalizer then placed his hands under his legs (i.e., sat on his hands). At this point, the experimenter began a timer allotting 20 seconds for the turn. The vocalizer was allowed to produce as many sounds during this time as he wished. The turn ended either when a correct guess was made, or when the experimenter interrupted play to announce that the 20 seconds had expired. In this latter case, the vocalizer stated the correct word, and then moved on to the next turn. Players switched roles between each 15-word stack, so that the game was played in four sets. The vocalizer wore a wireless lapel microphone that was connected to a digital recorder, which recorded the experiment as an uncompressed .wav file at a sampling rate of 44.1 kHz.

Analysis

Analysis focused only on the vocalizations that were produced during the game.[2] Each sound was first identified as an attempted sound or as commentary to be discarded from further analysis (e.g., "uh"s and "um"s to express production difficulty, or commentary like "Man, this is a hard one"). Praat phonetic analysis software (Boersma, 2001) was used to measure each sound along six acoustic properties: *mean pitch* (fundamental frequency (F0) in Hertz), *pitch change* (the ordered difference between the maximum and minimum F0 in Hertz), *pitch range* (the absolute value of the difference between the maximum and minimum F0 in Hertz), *duration* (in seconds), *intensity* (the energy or loudness of the sound in decibels), and *harmonics-to-noise ratio* (HNR; a ratio of periodic to non-periodic energy in a sound, measured in decibels, with higher values reflecting a more tonal and less hoarse quality). We initially observed that two pairs of meanings — *fast*|*slow* and *many*|*few* — were often expressed by the repetition of relatively short sounds. Thus, for these meanings, an additional measure of *repetition rate* (repetitions per second) was included in place of the pitch change and pitch range measurements, which often could not be adequately measured because of the short duration of many of the sounds. For this measurement, each repetition of a local maximum in intensity was counted as a sound, even when separate sounds were not fully distinct. The boundary between sounds was the point of minimum energy between two local maxima.

For statistical analyses, multiple sounds produced in a single turn were averaged together for the six acoustic measures, and repetition rate was measured through the whole turn. Thus each participant provided one averaged token for each of their 30 words. Statistical tests consisted of a series of paired samples *t*-tests that compared antonyms along each of the acoustic measures.

Results

In total, 28 pairs of antonymic meanings were compared along 6 variables, and 2 pairs were compared along 5 variables, for a total of 178 statistical tests. Table 1 presents the significant results for each pair at a range of alpha levels. 26 pairs showed a reliable difference along at least one variable at $\alpha = .05$. (Tables with the complete results are available at http://mperlman.org/vocalcharades, which also

[2] The guesses participants made are, of course, interesting too, but their analysis is beyond the scope of the present paper. Guesses appeared to reflect a degree of semantic similarity to the correct answer, but the proportion of exactly correct responses was low due to the open-ended nature of the task.

includes exemplary sounds for each meaning.) To account for the likelihood of type I error, a Holm-Bonferroni correction was applied to the series of tests performed on each meaning pair. By this approach, the most statistically significant measurement of a pair was compared to $\alpha = .05$ divided by the number of tests k (e.g., $k = 6$, $\alpha = .05\ /\ 6 = .0083$). If this criterion for significance was met, then the next most reliable effect was compared to α divided by $k-1$ ($\alpha = 0.05/5 = 0.01$), and so on, until the comparison no longer reached significance. After correction, 20 of the antonym pairs (66.7%) showed a reliable difference in at least one acoustic property.

These 20 pairs of antonyms were reliably distinguished by 12 unique combinations of variables (e.g., *fast*|*slow* by mean pitch and intensity vs. *alive*|*dead* by

Table 1. Distinguishing characteristics for each word pair.

Word pair	Primary ($p < 0.001$)	Strong ($p < 0.01$)	Moderate ($p < 0.05$)	Marginal ($p < 0.1$)
Alive – Dead	**Intensity ←** **HNR ←**			Pitch range ←
Antagonistic – Friendly		**HNR →**		
Attractive – Ugly			Mean pitch ← Pitch range ← HNR ←	
Bad – Good	**HNR →** **Pitch range →**	**Mean pitch →** **Duration →**		
Big – Small		**Mean pitch →** **HNR →** **Intensity ←**	**Duration ←**	
Bright – Dark		**Mean pitch ←**	Intensity ← Pitch change ←	Pitch range ←
Cold – Hot		**HNR ←**	Mean pitch → Duration ← Pitch range →	
Difficult – Easy	**Duration ←**			Mean pitch →
Down – Up	**Pitch change →**			Mean pitch →
Dry – Wet				Intensity →
Dull – Sharp		**Mean pitch →**		Intensity →
Fast – Slow		**Mean pitch ←**	**Intensity ←** HNR → Rep. rate ← Duration →	

Table 1. (*continued*)

Word pair	**Primary** **($p < 0.001$)**	**Strong** **($p < 0.01$)**	**Moderate** **($p < 0.05$)**	**Marginal** **($p < 0.1$)**
Female – Male	**Mean pitch ←**	**HNR ←**		Pitch range ←
Few – Many		**Rep. rate →**		
Hard – Soft			Duration →	
Heavy – Light (weight)				
Here – There			Duration →	Pitch range → Pitch change ←
Last year – Next year				HNR ←
Lift up – Set down			Duration ← Pitch range ← Mean pitch ← Pitch change ←	
Long – Short	**Duration ←**			Mean pitch →
New – Old		**HNR ←** **Intensity ←**	Duration →	Mean pitch ← Pitch range ← Pitch change →
No – Yes		**Mean pitch →**	Pitch range → HNR →	
Now – Later		**Duration →** **Intensity ←** **HNR →**		
Nutritious – Poisonous	**HNR ←**		Mean pitch → Intensity ←	
Predator – Prey			HNR →	
Rough – Smooth	**HNR →**		Pitch change ← Pitch range → Mean pitch →	
Start – Stop				
Straight – To-the-side			Pitch range → HNR ←	
Strong – Weak	**Intensity ←**		Mean pitch →	
Surprising – Predictable	**Mean pitch ←**	**Intensity ←**	Pitch range ← HNR →	Duration →

Note: The left or right arrows indicate that the left or right item of the pair has the higher value. Bold indicates that the word pair or characteristic was significant after Holm-Bonferroni correction.

intensity and HNR). Of the six standard measurements, harmonics-to-noise ratio (53% of pairs) and mean pitch (47%) most frequently distinguished antonymic words (without correction; see Supplementary Tables S1 and S2). The next most frequently significant variables were duration (37%), pitch range (29%), intensity (29%), and pitch change (14%). Repetition rate was significant in distinguishing both tested pairs of words.

Discussion

Language and gesture scholars have often assumed that vocalizations are one-dimensional, and thus, are extremely limited in their potential for iconic expression, especially in comparison to manual gestures. We tested this assumption by examining whether people are able to use prosodic features of their voice for the iconic expression of different kinds of meanings. Participants played a vocal charades game in which they attempted to communicate 60 different meanings that included 30 pairs of antonyms. Under conservative evaluation, they reliably produced similar vocalizations that distinguished the antonyms of 20 pairs, and as many as 26 pairs under less conservative assessment. These results illustrate the consistent ways that people feel qualities of their voice are expressive of different meanings, and arguably reflect iconic mappings between sound and meaning.

As mentioned above, participants could see each other and very likely could not help but express themselves with visible, bodily movement. However, this point does not account for the high degree of consistency across participants in the acoustic properties of the vocalizations they produced for different meanings. Indeed, most experiments to date relate to the ability to interpret the meanings of iconic sounds and vocalizations; it is the ability to produce iconic vocalizations that remains much more in question.

As we discuss in more detail below, many of the meanings for which participants produced iconic vocalizations figure prominently into important semantic and grammatical distinctions that are common across languages. The range of meanings included adjectives relating to animate (e.g., *strong*|*weak*) and inanimate entities (*cold*|*hot*), space (*up*|*down*), texture (*rough*|*smooth*), information status (*surprising*|*predictable*), pragmatic functions (*no*|*yes*), time (*now*|*later*), physical extension and size (*big*|*small*), number (*few*|*many*), manner of motion (*fast*|*slow*), gender (*female*|*male*), luminance (*bright*|*dark*), animacy (*alive*|*dead*), social intention (*friendly*|*antagonistic*), and emotional valence (*bad*|*good*). These results demonstrate the rich potential of iconicity in the vocal modality.

Iconicity and convention

Our claim is that the consistent acoustic properties that participants generated in their vocalizations for different meanings reflect iconic relationships between form and meaning. However, one qualification that deserves discussion relates to the possible influence of participants' shared cultural backgrounds. Although players were not permitted to use words, they may still have made use of culturally mediated mappings between nonlinguistic sounds and meaning. This raises the possibility that the similarity of the sounds participants produced might have resulted from common arbitrary convention, rather than a sense of correspondence between form and meaning.

The role of culture is important to consider and warrants future experiments with speakers of different linguistic and cultural backgrounds. Nonetheless, as is generally the case in systems of gesture and language, we note that iconicity and convention are orthogonal qualities, illustrated, for example, by the widespread persistence of iconicity in mature, conventionalized signed languages (Armstrong & Wilcox, 2007; Taub, 2001). While the establishment of convention emancipates a gesture from iconicity and thus frees it to develop towards a more arbitrary form, the symbol can nevertheless retain its iconicity.

An example of this in the vocal modality is onomatopoetic words for animal sounds, which vary substantially from language to language, but are still quite constrained by iconicity. As Violi (2000) points out, /psss/ could not pass for the sound of a cockerel. An example from the present study is the use by some participants of a "wolf whistle" to express *attractive*. While this particular form is certainly learned, its acoustic properties of high pitch and harmonicity plausibly originate from more general iconic associations with the meaning. For instance, the high pitch may be associated with qualities like positive arousal such as with pleasant surprise, and additionally, sounds with purer tones may be perceived as more attractive than aperiodic sounds (cf. a deep growl).

Modality and the nature of vocalization: Comparisons with manual gesture

Hockett (1978) noted that life is four-dimensional in space and time, and reasoned that vocalization must therefore force the multidimensional aspects of life that a speaker wishes to communicate into a single dimension. In contrast, he remarks that the dimensionality of gesturing "is that of life itself" (p. 274), and as such, the gestural modality enables a more natural and direct representation of a speaker's meaning (also see, e.g., Armstrong & Wilcox, 2007, pp. 113–114). Thus far, we have argued against this stark qualitative disparity between gesture and vocalization. Instead, we suggest that a more nuanced comparison between the modalities is needed.

To begin with, this comparison raises the question of what we mean by the notion of *modality* in the first place. Are we referring to the articulators, for example, the hands versus the vocal tract, and their capabilities? Or to whether the medium of transmission is optical or acoustic and perceived by the eyes or by the ears? An additional consideration is that the common emphasis on manual gesturing as communicating exclusively through visual information, and vocalizations through acoustic information is not accurate. Manual gestures can be audible (e.g., clapping, snapping, and slapping — a fact that gains more attention in studies of ape gesture, e.g., Call & Tomasello, 2007), and the visible components of speaking, including its associated facial expressions, are known to be very important in face-to-face communication (Massaro, 1998).

From an articulatory perspective, it appears possible that vocalization exhibits a comparable degree of life-like dimensionality to manual gesturing. Like manual gesture, vocalizations involve the complex coordination of multiple articulators that operate over the three dimensions of space plus time. This dimensionality and its potential for iconicity is illustrated by our vocal charades results, where the dimensions of vocalization were analyzed in terms of the array of acoustic variables that served in the expression of different meanings. The variety of unique acoustic signatures characterizing the differences between antonymic meanings indicates the potential for each variable to act independently and meaningfully from the others. Moreover, even these seven measurements are quite coarse when one considers the potential for expression through more finely grained phonetic and prosodic qualities.

Thus we propose that instead of claiming that one modality is much better suited for iconic expression than the other, a more fruitful approach is to examine the detailed ways in which each modality can function in the iconic expression of particular kinds of meanings. Manual gesture is likely better suited for some domains of iconic expression, and vocalization for others. Furthermore, the ability to coordinate iconic expressions across modalities may be especially useful, for it may serve to increase the robustness of communication across different environmental conditions (e.g., background noise, low lighting, etc.), or to communicate different kinds of information. By considering the ways that gestural and vocal communicative actions can complement and supplement each other, we can develop a more detailed framework to understand how languages may have evolved over human history. In the discussion that follows, we consider some important domains of meaning and communicative functions that are common in human communication, and assess the suitability of iconic manual gesture and vocalization for their expression.

We have described above how many scholars have observed that gesture very naturally affords iconic expression. Kendon (2014) explains that much of this semiotic potential stems from the performance of "manipulatory actions acting in a virtual world" (p. 6). Gestures naturally resemble the kinematics of real actions, and thereby afford the relatively direct representation of actions, particularly from a character viewpoint (Cartmill, Beilock, & Goldin-Meadow, 2012). As gestures evoke real-world manipulatory actions, they can also refer to objects that are associated with those actions. In addition, precise shapes are easily depicted by using an extended index finger to virtually trace the outline of objects in space, and likewise the planes of one's hands can trace the contour of a surface or the shape of an object. Manner and path of movement can be expressed by similar means. In comparison to gesture, the resemblance between vocalization and manipulatory action is narrower in scope, with direct resemblance limited to vocal actions, such as the production of speech or other vocalizations. Notably, this resemblance is incorporated into spoken languages via the grammatical device of quotation.

Another iconic affordance of gesture is the use of the hands themselves to represent objects, which enables the gesturer to depict actions from an observer viewpoint (Cartmill et al., 2012). The virtue of this capacity is demonstrated by Armstrong, Stokoe and Wilcox (1995, p. 22) in their discussion of the American Sign Language sign meaning to *capture* or *grasp*. To articulate the sign, one hand reaches and grasps the upright forefinger of the other hand. The authors point out the natural way this action depicts a transitive construction in which an agent (the grasping hand) acts upon a patient (the grasped forefinger). This sort of iconic expression is argued to be the archetype of the original kinds of iconic gestures that might have supported more complex grammatical constructions of human language. In contrast, vocalization does not offer this same iconic means to represent such a construction. The vocal tract does not afford the simultaneous representation of multiple entities (as does a parallel set of hands), and consequently it must represent transitive events in serial order.

Related to the act of reaching, gestures also enable a person to point at something out in the world to establish joint attention on that thing with an interlocutor (Tomasello, 2008). By extending one's arm, hand, and fingers into a vector, a person can direct another's attention in a precise way that is not possible with one's voice. Building on this natural capacity for deixis, gesture further provides an effective means to depict spatial layouts by pointing and virtual placement. This is clearly evident in signed conversation, where absent objects are virtually "placed" at different locations in front of the speaker and referenced by pointing to their respective locations (Emmorey, 2002; Liddell, 2003). However, while vocalizations are generally ill suited for pointing, they do afford a punctual means to "point" to oneself (i.e., *me*, *here*) and to the present time (i.e., *now*). Clark (2003) gives the example of communicating one's location in the dark, but the significance of the ability to draw attention to oneself or to mark a particular moment in time is broader. The main advantage of vocalization over gesture in this case is that one can point to the here and now without the visual attention of the interlocutor. (And, of course, more generally, the fact that vocalizations are perceptible without visual attention is a quite significant point.)

Exceptions aside, it is evident that in these domains of expression, gestures are supremely well suited for iconic representation of visible properties and patterns of action. However, this does not imply, as Tomasello (2008) concludes, for example, that people are completely unable to communicate about these domains through iconic vocalizations. Even if secondary, vocalizations have the potential to enrich gestural representations with reinforcing or supplementary information. Many actions and motions are associated with characteristic sounds (e.g., knocking on a door), and the iconic representations of those sounds can convey rich details about action and motion (e.g., the intensity of knocking or the hardness of the door). (For a sense of the potential breadth of this domain, see the many hundreds of listings in Taylor's (2007) dictionary of comic book words.) Indeed, the present results highlight ways in which vocalizations might serve to emphasize qualities like manner of motion (e.g., *fast*, *slow*, *difficult*, *easy*) or physical properties about a relevant object that is virtually acted upon (e.g., *big*, *small*, *rough*, *smooth*, *hard*, *soft*). In this respect, it is interesting to note that expressive words across languages characteristically bear adjectival and adverbial meanings (Nuckolls, 2004).

The coordination of iconic vocalization with gestures is illustrated in video of speaker George Lindell, who describes his experience of an immediately recent car accident (https://web.archive.org/web/20151030221906/https://www.youtube.com/watch?v=gP5mWgbgCEc). Lindell uses a variety of vocal sounds to punctuate the exciting moments of the event, such as the loud /bæm/ for the crash and /fum fum/ for electric sparks flying through the air. His animated communication shows how vocalization may be especially well suited for the expression of temporal qualities of actions, including aspectual properties like repetition and punctuality.

In similar ways, iconic vocalizations can enrich pointing and spatial gestures, helping to specify the target and the reason for pointing at it. This potential is suggested by findings across languages that proximal versus distal distinctions, including first versus second person, tend to exhibit similar distinctive phonetic patterns (Tanz, 1971; Wichmann, 2010). Our results show further how vocalization could supply additional information about proximal (e.g., *here*, *now*, *near*, *short*) and distal (*there*, *later*, *far*, *long*) locations in space and time.

We also suggest that there are domains in which vocalization excels in iconic expression. One such domain is the expression of emotion (Banse & Scherer, 1996; Cosmides, 1983; Sauter, Eisner, Ekman, & Scott, 2010). The use of prosody for the iconic expression of emotion is understood to be motivated by a deeply natural connection, and consequently, it is often disregarded as a trivial case of vocal iconicity (e.g., Shintel et al., 2006). Bolinger (1986), for example, proposed that intonation originated as "part of a gestural complex whose primitive and still surviving function is the signaling of emotion" (p. 195). For instance, vocalizations with high pitch and intensity are naturally associated with excitement and high arousal, while low-pitch, low-intensity vocalizations indicate low arousal and boredom. Moreover, the articulation of different vowels can involve different facial muscles for smiling (/i:/) and frowning (/o:/), and pronouncing these vowels influences one's emotional state (Rummer, Schweppe, Schlegelmilch, & Grice, 2014).

However, the primitive origins of the relationship between voice and emotion do not entail that it is narrow or trivial. For example, Bolinger suggests that this natural mapping between voice and emotion motivates the use of intonation to accent new versus old information, a critical distinction in communication that he notes children begin to produce from a young age. The results of the present study further demonstrate the reach of vocalization to emotionally laden meanings such as *good* | *bad*, *no* | *yes*, and *attractive* | *ugly*, as well as information status like *predictable* | *surprising* and *new* | *old*.

Vocalization also appears apt for the expression of gender and size. The rather obvious importance of gender in human experience is reflected in the pervasive extent to which it figures in the lexicons and grammars of spoken languages (Corbett, 1994). Our results show how the concepts of *female* and *male* are consistently expressed by high and low pitch, a pattern that is also seen in many languages, as for example in the second formant resonance of the English feminine suffix /i/. This pattern is also found in English personal names, as vowels like (/i/, /e/) are associated more frequently with female names and vowels like (/a/, /o/) with male names (Pitcher, Mesoudi, & McElligott, 2013). The same English suffix also has a diminutive meaning — a pattern common across languages (Ultan, 1978) — and likewise, we found that *small* is expressed with higher pitch than *large*.

More broadly, Ohala (1994) argues that the natural expression of size through iconic pitch is the basis for a number of important functions of intonation in speech, including the marking of questions, and also the expression of affective qualities like *deference* vs. *assertiveness*, *politeness* vs. *authority*, *submission* vs. *aggression*, and *confidence* vs. *lack of confidence*. He suggests that the association between size and pitch originates from the physiology of tetrapod vertebrate vocal tracts — larger animals tend to have larger vocal tracts that produce lower pitched sounds. This distinction appears to be ritualized in the voices of male humans compared to females and children. More generally, experiential correlations between pitch and size, vocal or otherwise, may reinforce their association.

Another strength of vocalization for iconic expression stems from the ease of extending vocalizations through time and the ability to articulate sharp boundaries between sounds. Above we pointed out how this allows one to supplement gestures about actions and events with temporal, aspectual information. Our results also demonstrate how vocalization can express plurality, using the production of more sounds to express *many* compared to fewer sounds for *few*. This reflects the iconic morphological principle that plurality is typically marked in languages and expressed with additional or longer morphology (Wescott, 1971). The expression of plurality through the repetition of sound also forms the iconic basis for the grammatical device of reduplication, which is especially associated with ideophones and similar iconic lexical classes (Dingemanse, 2012). Repetition can also be used to express a recurring element of an action or event. This is illustrated in an overheard example of a hunter recounting an incident in which he climbed a tall ladder into a deer stand (and fell off). Reenacting the story in dramatic fashion, he conveyed the extent of the high climb by repetition: "Climb climb climb climb climb", helping the listener to imagine each rung of the way.

Finally, the most obvious strength of vocalizations for iconic representation would seem to be the imitation of sound (lexicalized in onomatopoeia). In industrial civilizations, this might include the vast array of mechanical and electronic noises, as well as all of the noises of the humans (vocal and otherwise), birds and other animals that surround us. The significance of communicating through the iconic representation of sounds is often played down, leading many to regard its scope as quite limited (hence the disparagingly labeled "bow-wow" or "flatulence" theories of language origins). However, one might take a moment to listen and observe that a large swath of human experiences and concepts are associated with sound.

It is worth noting that the strength of the human capacity for vocal imitation is not obvious to all. For example, Pinker and Jackendoff propose (2005, p. 209):

> Humans are not notably talented at vocal imitation in general, only at imitating speech sounds (and perhaps melodies). For example most humans lack the ability (found in some birds) to convincingly reproduce environmental sounds … Thus 'capacity for vocal imitation' in humans might better be described as a capacity to learn to produce speech.

Contrary to this position, a browse through YouTube reveals a number of cases of people who are quite skilled at vocal imitation, such as in beat boxing (e.g., https://web.archive.org/web/20151030223319/https://www.youtube.com/watch?), imitating animal sounds (e.g., https://web.archive.org/web/20151030224508/https://www.youtube.com/watch?) or car sounds (e.g., https://web.archive.org/web/20151030224925/https://www.youtube.com/watch?v=IF5rMU5rDXE). While Pinker and Jackendoff concede that some humans might have an exceptional capacity for the skill, ethnographic work suggests that humans who practice vocal imitation are generally able to develop it to a high degree of competence. For example, Lewis (2009) discusses the importance of the vocal mimicry of sound by the Mbendjele Pygmies living in the Congo. When describing an encounter with a dangerous animal in the jungle, Mbendjele speakers pay great attention to imitating the acoustic details of the event, including both sounds of the animal and also inanimate sounds like the thrashing of trees. The Mbendjele also utilize vocal imitation for hunting. By producing highly realistic imitations of an animal's play or mating calls, hunters can lure various species within range, from duikers to crocodiles.

Additionally, we suggest that Pinker and Jackendoff's bar of high fidelity imitation is arguably not the relevant standard for gauging the human capacity for vocal iconicity. Gestures are characteristically schematic rather than precise imitations, and the critical question is whether people are able to produce iconic vocalizations that are sufficiently accurate to be effective. They do not need to be lyre bird-like virtuosos (cf. https://web.archive.org/web/20151030225440/https://www.youtube.com/watch?v=VjE0Kdfos4Y). Indeed there is plenty of evidence across the lexicons of languages indicating the importance of the iconic representation of sounds from this more pragmatic standpoint. In English, for example, Rhodes (1994) documents a repertoire of over 100 words used to express what he describes as aural images — "click", "blabber", "hoot", "ring", "splash", "whoosh", etc. And as Oswalt (1994) points out, English sound words include a large repertoire for inanimate sounds as well as animate ones.

Thus we have illustrated how manual gestures may be better suited for the iconic expression of particular kinds of meanings, whereas vocalizations may be better for others. We have also suggested ways in which these modalities may work together to enrich communication and increase its robustness. Indeed, many of the examples that we have pointed out hint that multimodal iconic expression is often the case. However, it is clear that further research is needed on this topic. Although there has been a substantial amount of work showing the tight and complex relationship between speech and manual gesturing (Kendon, 2004; McNeill, 1992), very little research has investigated the coordination of iconic expression across modalities.

Iconicity and gesture in language evolution

A long-standing debate in language evolution concerns the question of whether languages evolved originally as gestures or as vocalizations. Many theories of language evolution have maintained that vocalizations must have been "bootstrapped" (Fay et al., 2013), or "piggybacked" (Tomasello, 2008, p. 330) on gestures "because of the vastly greater possibility for iconic productivity in the visual medium" (Armstrong & Wilcox, 2007, p. 123; see also, e.g., Corballis, 2002; Hewes, 1973). At some point in human evolution, people began to produce arbitrary vocalizations to accompany inherently meaningful iconic gestures and pointing, and through association, these vocalizations came to be independently meaningful as they conventionalized into the forms of speech. Tomasello suggests that the human capacity for language evolved mostly in the gestural modality, and that the use of vocalization may be a very recent development. Likewise, Corballis (2002) proposes a transition to speech at roughly 50,0000 years ago, while suggesting that gestural language may be as many as 1 million years old.

Counter to this point of view, the present work — along with the growing body of evidence we have reviewed — suggests that humans are, by their nature, highly inclined to embed iconic form-meaning resemblances in vocalizations. Thus as people have a proclivity to embody meaning iconically with their hands, they also appear to have a comparable proclivity to embody meaning with their voice. The ways in which vocalizations are iconic might be different from gestures, but they are not less rich or profound. Therefore, it is plausible to consider that iconic vocalizations may have played a significant role in the evolution of language.

This does not, however, lead to the conclusion that language sprung up exclusively in the vocal modality. Compelling evidence indicates that modern languages are flexible, highly coordinated, multimodal communication systems that involve elements of manual gesturing, posturing, and facial expression, in addition to vocalization (Kendon, 2004; McNeill, 1992). Kendon (2009) describes the natural state of spoken language as a speech-kinesis ensemble, pointing out that language origin theories must explain the tight coordination between speaking and gesturing (also see McNeill, 2012). He proposes that the "polymodalic" nature of communication has been in place for the history of the line leading to *Homo*, present originally as a vestige of the hand-mouth coordination that primates evolved for feeding (MacNeilage, 2008). He further suggests that the critical adaptation for languages was the ability to transform polymodal ensembles of praxic activity into a 'virtual' or 'as if' mode — that is, essentially, an iconic mode — that enabled the representation of meaning (Kendon, 1991, 2009; also see Perlman, Tanner, & King, 2012b). However, missing from Kendon's account is serious consideration of how "as if" modes of thought might be extended to vocalizations. His explicit concern with the ensemble of kinesis and *speech* implies that the vocal modality is tantamount to the linguistic channel. We propose instead that the more aptly balanced ensemble is kinesis and *vocalization*.

This proposal presents the challenge to explain how the capacity for vocal iconicity has evolved over the course of hominid evolution, including its relationship with manual gesture. Relevant to this question are the relative gestural and vocal abilities of the great apes. Studies of ape communication have discovered that their gesturing is highly flexible, under voluntary control, and can involve considerable social learning (Call & Tomasello, 2007). In contrast, it was traditionally believed that apes lack voluntary control over their vocalization, as well as any capacity for vocal learning. This comparison is often presented as a primary piece of evidence in favor of a gestural origin of language.

However, increasingly, evidence shows that apes have considerably more voluntary control over their vocal apparatus and breathing than previously realized. For example, one report describes the pragmatic use of voluntarily produced, novel vocalizations by captive chimpanzees to gain the attention of a human caregiver who lacked a line of sight to their actions (Hopkins, Taglialatela, & Leavens, 2007). Another study reports a zoo-housed orangutan able to whistle and even imitate the number and duration of a keeper's whistle (Wich et al., 2009). Studies of free-ranging orangutans describe their production of learned oral sounds, including a lip sputter associated with nest building, and a kiss-squeak — a call of distress that is created by a sharp intake of air through pursed lips (Hardus, Lameira, Van Schaik, & Wich, 2009; Van Schaik et al., 2003). Interestingly, orangutans sometimes produce the kiss squeak while holding a stripped leaf to their mouth, which functions to decrease the fundamental pitch of the sound (Hardus et al., 2009). These modified calls, in particular, tend to be produced by smaller individuals in high distress, suggesting that they produce this culturally learned behavior to sound bigger and ward off predators. Thus this instance of novel oral sound production may even be seen to qualify as a rudimentary "as if" sort of behavior, of the type Kendon (1991) postulates is crucial to the evolution of language.

With intensive human interaction, apes may even develop more enhanced vocal capabilities. For example, the enculturated gorilla Koko has acquired a repertoire of roughly a dozen learned and voluntarily controlled vocal and other breathing-related behaviors, such as grunting into a telephone and huffing on the lenses of eyeglasses (Perlman & Clark, 2015; de Boer & Perlman, 2014; Perlman, Patterson, & Cohn, 2012a). Considering the evolution of a kinesis-vocalization ensemble, Perlman and Clark describe how with just a few exceptions, these behaviors tend to be performed in association with manual gestures (e.g., blowing forcefully into a flat palm brought to the mouth) or incorporated into manual action routines (e.g., vocalizing into a telephone). Perlman et al. additionally describe Koko's differential use of the intensity of exhalation in two different social actions — gentle breaths in a friendly greeting ritual, compared to a single forceful breath in a reprimanding action. While the precise origins of these behaviors are unknown, they are learned and not typical of gorillas. Thus they present two more examples of the ape capacity to learn behavior involving oral/vocal sound production with an iconic relationship between communicative function and form.

Thus the earliest hominids probably had some very rudimentary capacity for iconic expression through the vocal tract, which points to a rather ancient history for iconic vocalization. Nevertheless, great apes do appear to exhibit more advanced abilities for iconic expression with their hands compared to their voice (Perlman & Gibbs, 2013; Perlman, Tanner & King, 2012b). It therefore reasonable to consider that the relative load communicated by each modality is likely to have shifted over time, with gestures occupying the greater share early on. An evolving faculty for vocal iconicity — such as increasing capacities for vocal imitation and for the conception of cross-modal correspondences between sound and meaning — may have facilitated transfer of communicative load from the manual to the vocal articulators.

We propose that this transfer across modalities is likely to have been a dynamic process, rather than a monolithic, steady transfer of communication from hands to mouth. Discussion in the previous section shows how the transfer may have proceeded differently across different kinds of meanings as they vary in their affordance of iconic representation through gestural and vocal actions. The different communicative demands posed by particular ecological environments may also have played a role. In this respect, it is interesting to consider ethnographic studies documenting the cultural significance of vocal iconicity in the speech and ritual of some indigenous groups, such as the Kaluli in the rainforest of Papua New Guinea (Feld, 1996) or speakers of Pastaza Quechua living in the Amazonian rainforest of Ecuador (Nuckolls, 1996). Both groups live in environments that are dense in vegetation, resulting in poor visibility, and rich in sound — two factors that may contribute to the enhanced importance of sound-based iconicity in their cultures.

Such examples point to the possibility that over evolution and into the modern day, the qualities of language systems, such as their degree of reliance on one modality or the other, or the degree to which they incorporate iconicity across modalities, may adapt to the pressures of the particular semantic, sociocultural and ecological niches in which the systems are used (cf. the linguistic niche hypothesis, Lupyan & Dale, 2015). At the very least, we suggest, based on the present results and the accumulation of many other relevant findings, that theories of language evolution will need to account for certain facts that are becoming increasingly established about modern languages: namely, that they are multimodal systems that incorporate iconicity across modalities.

Conclusion

Kendon (2004) has remarked that utterances "may be constructed from speech or from visible bodily action or from combinations of these two modalities" (p. 7). Yet the present work and literature review show that it is important to be clear that the term "speech" must include more than just language in its vocal form. Just as visible actions that count as the kinesic component of an utterance span the range from arbitrary to iconic, so do the audible actions of an utterance's vocal component. The present work adds to a substantial body of evidence showing that vocalizations deemed iconic are far from limited to the expression of just onomatopoeia or emotions. Rather, people are able to use iconic vocalizations to express a variety of meanings that rivals the richness of manual gestures, although, as we have seen, the particular domains of meaning that can be represented iconically with each modality is different. Thus, scholarship concerned with human communication and its evolution might profit from considering iconicity, not just as we see it in visible bodily actions, but as we find it in audible actions, as well.

Acknowledgements

We are very grateful for several rounds of challenging critiques and suggestions by the editor, Adam Kendon. We are also grateful for helpful comments by an anonymous reviewer, as well as for stimulating discussion with Nate Clark, Ray Gibbs, Gary Lupyan, Dom Massaro, Lynn Perry, Joanne Tanner, Sherman Wilcox, Meg Wilson, Bodo Winter, and Adrienne Zihlman.

References

Ahlner, Felix & Jordan Zlatev (2011). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38, 298–348.

Alpher, Barry (2001). Ideophones in interaction with intonation and the expression of new information in some indigenous languages of Australia. In Erhard F. K. Voeltz & Christa Kilian-Hatz (Eds.), *Ideophones* (pp. 9–24). Amsterdam: John Benjamins. DOI: 10.1075/tsl.44.03alp

Arbib, Michael A. (2012). *How the brain got language*. Oxford: Oxford University Press. DOI: 10.1093/acprof:osobl/9780199896684.001.0001

Armstrong, David F., William C. Stokoe, & Sherman E. Wilcox (1995). *Gesture and the nature of language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620911

Armstrong, David F. & Sherman E. Wilcox (2007). *The gestural origin of language*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780195163483.001.0001

Banse, Rainer & Klaus R. Scherer (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636. DOI: 10.1037/0022-3514.70.3.614

Bentley, Madison & Edith J. Varon (1933). An accessory study of phonetic symbolism. *American Journal of Psychology*, 45, 76–86. DOI: 10.2307/1414187

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5, 341–345.

Bolinger, Dwight (1986). *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.

Bremner, Andrew J., Serge Caparos, Jules Davidoff, Jan de Fockert, Karina J. Linnell, & Charles Spence (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, Roger W., Abraham H. Black, & Arnold E. Horowitz (1955). Phonetic symbolism in natural languages. *Journal of Abnormal Social Psychology*, 50, 388–393. DOI: 10.1037/h0046820

Call, Josep & Michael Tomasello (Eds.) (2007). *The gestural communication of apes and monkeys*. London: Lawrence Erlbaum.

Cartmill, Erica A., Sian Beilock, & Susan Goldin-Meadow (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B*, 367, 129–143. DOI: 10.1098/rstb.2011.0162

Childs, G. Tucker (1994). African ideophones. In Leanne Hinton, Johanna Nichols, & John J. Ohala (Eds.), *Sound symbolism* (pp. 178–206). Cambridge: Cambridge University Press.

Clark, Herbert H. (2003). Pointing and placing. In Sotaro Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243–268). Hillsdale, NJ: Erlbaum.

Clark, Nathaniel, Marcus Perlman, & Marlene Johansson Falck (2014). The iconic use of pitch to express vertical space. In Barbara Dancygier, Mike Borkent, & Jennifer Hinnell (Eds.), *Language and the creative mind* (pp. 393–410). Stanford: SCLI Publications.

Corballis, Michael (2002). *From hand to mouth: The origins of language*. Princeton: Princeton University Press.

Corbett, Greville (1994). Gender and gender systems. In R. Asher (Ed.), *The Encyclopedia of language and linguistics*, Vol. 3 (pp. 1347–1353). Oxford: Pergamon Press.

Cosmides, Leda (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology*, 9, 864–881.

Cuskley, Christine (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5, 39–62.

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

de Boer, Bart & Marcus Perlman (2014). Physical mechanisms may be as important as brain mechanisms in evolution of speech. *Behavioral and Brain Sciences*, 37 (6), 552–553.

Diffloth, Gerard (1972). The notes on expressive meaning. In Paul M. Peranteau, Judith N. Levi, & Gloria C. Phares (Eds.), *Papers from the Eighth Regional Meeting of Chicago Linguistic Society* (pp. 440–447). Chicago: Chicago Linguistic Society.

Dingemanse, Mark (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6, 654–672. DOI: 10.1002/lnc3.361

Dingemanse, Mark, Francisco Torreira, & N. J. Enfield (2013). Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE*, 8, e78273. DOI: 10.1371/journal.pone.0078273

Emmorey, Karen (2002). *Language, cognition, and brain: Insights from sign language research*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fay, Nicolas, Michael Arbib, & Simon Garrod (2013). How to bootstrap a human communication system. *Cognitive Science*, 37, 1356–1367. DOI: 10.1111/cogs.12048

Fay, Nicolas, Casey J. Lister, T. Mark Ellison, & Susan Goldin-Meadow (2014). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology*, 5, 1–12. DOI: 10.3389/fpsyg.2014.00354

Feld, Steven (1996). Waterfalls of song: an acoustemology of place resounding in Bosavi, Papua New Guinea. In Keith H. Basso & Steven Feld (Eds.), *Senses of Place* (pp. 91–135). Santa Fe, NM: School of American Research Advanced Seminar Series.

Gebels, Gustav (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning & Verbal Behavior*, 8, 310–312. DOI: 10.1016/S0022-5371(69)80083-6

Goldin-Meadow, Susan (2003). *The resilience of language: What gesture creation in deaf children can tell us about how children learn language*. New York, NY: Psychology Press.

Greenberg, Joseph H. (1966). Some universals of language with special reference to the order of meaningful constituents. In Joseph Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.

Greenberg, Joseph H. & James J. Jenkins (1966). Studies in the psychological correlates of the sound system of American English. *Word*, 22, 207–242. DOI: 10.1080/00437956.1966.11435451

Haiman, John (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, 56, 515–540. DOI: 10.2307/414448

Hamano, Shoko (1998). *The sound-symbolic system of Japanese*. Stanford, CA: CSLI.

Hardus, Madeleine E., Adriano R. Lameira, Carel P. Van Schaik, & Serge A. Wich (2009). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B*, 276, 3689–3694. DOI: 10.1098/rspb.2009.1027

Hewes, Gordon W. (1973). Primate communication and the gestural origins of language. *Current Anthropology*, 14, 5–24. DOI: 10.1086/201401

Hockett, Charles (1978). In search of Jove's brow. *American Speech*, 53, 243–315. DOI: 10.2307/455140

Hopkins, William D., Jared P. Taglialatela, & David A. Leavens (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behavior*, 73, 281–286. DOI: 10.1016/j.anbehav.2006.08.004

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

Imai, Mutsumi, Sotaro Kita, Miho Nagumo, & Hiroyuki Okada (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54–65. DOI: 10.1016/j.cognition.2008.07.015

Jakobson, Roman & Linda R. Waugh (1979). *The sound shape of language*. Bloomington, IN: Indiana University Press.

Jespersen, Otto (1933). Symbolic value of the vowel i. In Otto Jespersen (Ed.), *Linguistica* (pp. 283–303). Copenhagen: Levin & Munksgaard.

Kantartzis, Katerina, Mutsumi Imai, & Sotaro Kita (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, 35, 575–586. DOI: 10.1111/j.1551-6709.2010.01169.x

Kelly, Barbara F., William Leben, & Robert Cohen (2003). The meanings of consonants. *Proceedings of the 29th Berkeley Linguistics Society* (pp. 245–253).

Kendon, Adam (1991). Some considerations for a theory of language origins. *Man*, (N.S.) 26, 602–619. DOI: 10.2307/2803829

Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511807572

Kendon, Adam (2009). Language's matrix. *Gesture*, 9, 352–372. DOI: 10.1075/gest.9.3.05ken

Kendon, Adam (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130293 DOI: 10.1098/rstb.2013.0293

Klink, Richard R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, 11, 5–20. DOI: 10.1023/A:1008184423824

Köhler, Wolfgang (1929). *Gestalt Psychology*. New York: Liveright.

Kovic, Vanja, Kim Plunkett, & Gert Westermann (2010). The shape of words in the brain. *Cognition*, 114, 19–28. DOI: 10.1016/j.cognition.2009.08.016

LaPolla, Randy J. (1994). An investigation of phonetic symbolism as it relates to Mandarin Chinese. In Leanne Hinton, Johanna Nichols, & John J. Ohala (Eds.), *Sound symbolism*. Cambridge: Cambridge University Press.

Lewis, Jerome (2009). As well as words: Congo Pygmy hunting, mimicry, and play. In Rudolf Botha & Chris Knight (Eds.), *The cradle of language* (pp. 236–256). Oxford, UK: Oxford University Press.

Liddell, Scott K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511615054

Lupyan, Gary & Daniel Casasanto (2014). Meaningless words promote meaningful categorization. *Language and Cognition*, FirstView, 1–27.

Lupyan, Gary & Rick Dale (2015). The role of adaptation in understanding linguistic diversity. In Randy LaPolla & Rik De Busser (Eds.), *The shaping of language: The relationship between the structures of languages and their social, cultural, historical, and natural environments* (pp. 289–316).

MacNeilage, Peter F. (2008). *Origin of speech*. Oxford: Oxford University Press.

Massaro, Dominic W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Maurer, Daphne, Thanujeni Pathman, & Catherine J. Mondloch (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

McGregor, William (2001). Ideophones as the source of verb in northern Australian languages. In Erhard F. K. Voeltz & Christa Kilian-Hatz (Eds.), *Ideophones* (pp. 205–221). Amsterdam: John Benjamins. DOI: 10.1075/tsl.44.17mcg

McNeill, David (1992). *Hand and mind*. Chicago: University of Chicago Press.

McNeill, David (2012). *How language began: Gesture and speech in human evolution*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139108669

Mikone, Eve (2001). Ideophones in the Balto-Finnic languages. In Erhard F. K.Voeltz & Christa Kilian-Hatz (Eds.), *Ideophones* (pp. 223–233). Amsterdam: John Benjamins. DOI: 10.1075/tsl.44.18mik

Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen, & Simon Kirby (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130299. DOI: 10.1098/rstb.2013.0299

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

Nuckolls, Janis B. (1996). *Sounds like life: Sound-symbolic grammar, performance, and cognition in Pastaza Quechua*. New York: Oxford University Press.

Nuckolls, Janis B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28, 255–282. DOI: 10.1146/annurev.anthro.28.1.225

Nuckolls, Janis B. (2004). To be or not to be ideophonically impoverished. In Wai Fong Chiang, Elaine Chun, Laura Mahalingappa, & Siri Mehus (Eds.), *SALSA XI: Proceedings of the Eleventh Annual Symposium about Language and Society* (pp. 131–142). Austin: University of Texas.

Nygaard, Lynne, Debora Herold, & Laura Namy (2009). The semantics of prosody: Acoustic and perceptual correlates to word meaning. *Cognitive Science*, 33, 127–146. DOI: 10.1111/j.1551-6709.2008.01007.x

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

Oswalt, Robert (1994). Inanimate imitatives in English. In Leanne Hinton, Johanna Nichols, & John J. Ohala (Eds.), *Sound symbolism* (pp. 293–306). Cambridge: Cambridge University Press.

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

Peirce, Charles S. (1955) Logic as semiotic: The theory of signs. In Justus Buchler (Ed.), *Philosophical writings of Peirce* (pp. 99–119). New York: Dover.

Perlman, Marcus (2010). Talking fast: The use of speech rate as iconic gesture. In Fey Perrill, Mark Turner, & Vera Tobin (Eds.), *Meaning, form, and body* (pp. 245–262). Stanford: CSLI Publications.

Perlman, Marcus & Nathaniel Clark (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition*, 18, 1165–1179. DOI: 10.1007/s10071-015-0889-6

Perlman, Marcus, Francine G. Patterson, & Ronald H. Cohn (2012a). The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biolinguistics*, 6, 433–444.

Perlman, Marcus, Joanne E. Tanner, & Barbara J. King (2012b). A mother gorilla's variable use of touch to guide her infant: Insights into iconicity and the relationship between gesture and action. In Simona Pika & Katja Liebal (Eds.), *Developments in primate gesture research* (pp. 55–72). Amsterdam: John Benjamins. DOI: 10.1075/gs.6.04per

Perlman, Marcus & Raymond W. Gibbs Jr. (2013). Pantomimic gestures reveal the sensorimotor imagery of a human-fostered gorilla. *Journal of Mental Imagery*, 37, 73–96.

Perlman, Marcus, Rick Dale, & Gary Lupyan (2014). Iterative vocal charades: The emergence of conventions in vocal communication. In Erica A. Cartmill, Sean Roberts, Heidi Lyn, & Hannah Cornish (Eds.), *The evolution of language: Proceedings of the 10th International Conference* (*EVOLANG10*). New Jersey: World Scientific.

Perlman, Marcus, Nathaniel Clark, & Marlene Johansson Falck (2015). Iconic prosody in story reading. *Cognitive Science*, 39 (6), 1348–1368. DOI: 10.1111/cogs.12190

Perniss, Pamela & Gabriella Vigliocco (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130300. DOI: 10.1098/rstb.2014.0179

Pinker, Steven & Ray Jackendoff (2005). What's special about the human language faculty? *Cognition*, 95, 201–236. DOI: 10.1016/j.cognition.2004.08.004

Pitcher, Benjamin J., Alex Mesoudi, & Alan G. McElligott (2013). Sex-biased sound symbolism in English-language first names. *PLoS ONE*, 8, e64825. DOI: 10.1371/journal.pone.0064825

Ramachandran, Vilayanur S. & Edward M. Hubbard (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3–34.

Rhodes, Richard. Aural images. In Leanne Hinton, Johanna Nichols, & John J. Ohala (Eds.), *Sound symbolism* (pp. 276–292). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511751806.

Rummer, Ralf, Judith Schweppe, René Schlegelmilch, & Martine Grice (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion*, 14, 246–250. DOI: 10.1037/a0035752

Sandler, Wendy (2013). Vive la différence: Sign language and spoken language in language evolution. *Language and Cognition*, 5, 189–203. DOI: 10.1515/langcog-2013-0013

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

Saussure, Ferdinand de ([1916] 1966). *Course in general linguistics*. New York: McGraw-Hill.

Sauter, Disa A., Frank Eisner, Paul Ekman, & Sophie K. Scott (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107, 2408–2412.

Schultze-Berndt, Eva (2001). Ideophone-like characteristics of uninflected predicates in Jaminjung (Australia). In Erhard F. K.Voeltz & Christa Kilian-Hatz (Eds.), *Ideophones* (pp. 355–373). Amsterdam: John Benjamins. DOI: 10.1075/tsl.44.27sch

Shintel, Hadas & Howard C. Nusbaum (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, 105, 681–690. DOI: 10.1016/j.cognition.2006.11.005

Shintel, Hadas & Howard C. Nusbaum (2008). Moving to the speed of sound: Context modulation of the effect of acoustic properties of speech. *Cognitive Science*, 32, 1063–1074. DOI: 10.1080/03640210801897831

Shintel, Hadas, Howard C. Nusbaum, & Arika Okrent (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55, 167–177. DOI: 10.1016/j.jml.2006.03.002

Steklis, Horst D. & Stevan Harnad (1976). From hand to mouth: Some critical stages in the evolution of language. In Stevan Harnad, Horst D. Steklis, & Jane B. Lancaster (Eds.), *Origins and evolution of language and speech. Annals of the New York academy of sciences*, 280, 445–455.

Tanz, Christine (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, 14, 266–276.

Taub, Sarah (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511509629

Taylor, Kevin J. (2007). *KA-BOOM! A dictionary of comic book words, symbols & onomatopoeia*. USA: Lulu.com.

Thompson, Patrick D. & Zachary Estes (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, 64, 2392–2404. DOI: 10.1080/17470218.2011.605898

Tomasello, Michael (2008). *Origins of human communication*. Cambridge: MIT Press.

Ultan, R. (1978). Size-sound symbolism. In Joseph H. Greenberg (Ed.), *Universals of human language, Vol. 2: Phonology* (pp. 527–568). Stanford, CA: Stanford University Press.

Urban, Matthias (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

Van Schaik, Carel P., Marc Ancrenaz, Gwendolyn Borgen, Birute Galdikas, Cheryl D. Knott, Ian Singleton, Akira Suzuki, Sri Suci Utami, & Michelle Merrill (2003). Orangutan cultures and the evolution of material culture. *Science*, 3, 102–105. DOI: 10.1126/science.1078004

Violi, Patrizia (Ed.) (2000). *Phonosymbolism and poetic language*. Turnhout, Belgium: Brepols.

Waugh, Linda (2000). Against arbitrariness: Imitation and motivation revived. In Patrizia Violi (Ed.), *Phonosymbolism and poetic language* (pp.25–56). Turnhout, Belgium: Brepols.

Watson, Richard L. (2001). A comparison of some Southeast Asian ideophones with some African ideophones. In Erhard F. K.Voeltz & Christa Kilian-Hatz (Eds.), *Ideophones* (pp. 385–405). Amsterdam: John Benjamins. DOI: 10.1075/tsl.44.29wat

Westbury, Chris (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.

Wescott, Roger W. (1971). Linguistic iconism. *Language*, 47, 416–428. DOI: 10.2307/412089

Wichmann, Søren, Eric W. Holman, & Cecil H. Brown (2010). Sound symbolism in basic vocabulary. *Entropy*, 12, 844–858. DOI: 10.3390/e12040844

Corresponding author's address

Marcus Perlman
Department of Psychology
University of Wisconsin, Madison
1220 E. Johnson St.
Madison, WI 53703
USA

mperlman@wisc.edu

About the authors

Marcus Perlman is a postdoctoral Research Associate in the Department of Psychology at the University of Wisconsin-Madison. His research examines iconicity in speech and gesture, ape communication, and the evolution of language.

Ashley A. Cain received a B.A. from University of California, Santa Cruz, where she collaborated in cognitive psychology research related to language. Her projects focused on acoustic onset and vocal gesturing. Currently, she is a graduate student in the Experimental Psychology Masters program at San José State university, where she is applying psychology to cyber security. Her research interests include modeling of situation awareness, human-automation teamwork, and the relationship between trustworthiness and security.