Debunking two myths against vocal origins of language:

Language is iconic and multimodal to the core

Marcus Perlman

Language and Cognition Department

Max Planck Institute for Psycholinguistics

To appear in *Interaction Studies*

Send correspondence to:

Marcus Perlman

marcus.perlman@mpi.nl

Abstract

Gesture-first theories of language origins often raise two unsubstantiated arguments against vocal origins. First, they argue that great ape vocal behavior is highly constrained, limited to a fixed, species-typical repertoire of reflexive calls. Second, they argue that vocalizations lack any significant potential to ground meaning through iconicity, or resemblance between form and meaning. This paper reviews the considerable evidence that debunks these two "myths". Accumulating evidence shows that the great apes exercise voluntary control over their vocal behavior, including their breathing, larynx, and supralaryngeal articulators. They are also able to learn new vocal behaviors, and even show some rudimentary ability for vocal imitation. In addition, an abundance of research demonstrates that the vocal modality affords rich potential for iconicity. People can understand iconicity in sound symbolism, and they can produce iconic vocalizations to communicate a diverse range of meanings. Thus, two of the primary arguments against vocal origins theories are not tenable. As an alternative, the paper concludes that the origins of language – going as far back as our last common ancestor with great apes – are rooted in iconicity in both gesture and vocalization.

Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core

Introduction: Gestures versus vocalizations

Fundamental to theories of language evolution is the question of whether the first languages were spoken or signed. Although the vast majority of the roughly 7000 extant languages are characterized as spoken (Evans & Levinson, 2009), many scholars have argued that they must first have originated from the gestural communication of our hominid ancestors (Arbib, Liebal, & Pika, 2008; Armstrong & Wilcox, 2007; Corballis, 2003; Hewes, 1973; Hockett, 1978; Kendon, 1991; Levinson & Holler, 2014; Tomasello, 2008). According to this idea, gestures – prototypically visible, manual, communicative movements – were molded by generations of human ancestors into the first linguistic signing systems, perhaps reminiscent of how modern signed languages are created (Armstrong & Wilcox, 2007; Goldin-Meadow, 2016). Then, at some point in our history, gestural signs were replaced by vocalizations, and languages became predominantly spoken. In fact, some theories postulate that gesture was the dominant mode of human communication until relatively recently. For example, Tomasello (2008) suggested that symbolic communication might not have transferred to the vocal modality until as recently as 150,000 years ago, and Corballis (2002) proposed that the transition might even have occurred as recently as 50,000 years ago.

In large part, arguments for the gesture-first theory are based on *positive* evidence of the significance of gesture in human communication and in the communication of the great apes (e.g., Hewes, 1973; Kendon, 1991; Call & Tomasello, 2007). Gesturing is a

characteristic form of communication used by humans and great apes alike, connected to their manual dexterity and tool use (Cartmill, Beilock & Goldin-Meadow, 2012). For humans, gesture is used by people from all cultures and linguistic backgrounds (Feyereisen & de Lannoy, 1991). Children begin gesturing from an early age, often preceding the onset of spoken language (Bates, 1976), and even children with congenital blindness spontaneously use gestures (Iverson & Goldin-Meadow, 1998). Children also very naturally acquire signed languages when they are featured in the child's language-learning environment (Bonvillian, Orlansky, & Novack, 1983). And when deaf children lack a sense-suitable language model (i.e. a signed language), they readily improvise conventional home sign systems with their family members (Goldin-Meadow, 2003). In particular, they are able to improvise pointing and iconic gestures, which can ground understanding, and serve as the basis for negotiating more conventional signs. Over generations, interacting signers can develop fully grammatical and expressive signed languages (Sandler, Meir, Padden, Aronoff, 2005; Senghas, Kita, Ozuryek, 2004).

Gesturing is also prevalent in the communication of all four species of great apes: bonobos, chimpanzees, gorillas, and orangutans. Notably, their gesturing exhibits considerable social-cognitive sophistication, which may reflect homology with human gesture and language (Call & Tomasello, 2007; Tanner & Byrne, 1996). For example, apes use gestures intentionally, and adjust them to the attentional state of their audience (Leavens et al., 2010). Although the extent of learning is debated for the gesturing of wild apes (e.g. Graham, Furuichi, & Byrne, 2016; Hobaiter & Byrne, 2011), cases with human training show that their gestural repertoires are subject to extensive learning, reaching the order of a hundred gestures or more in a few cases (Fouts, 1997; Miles, 1990; Patterson,

& Linden, 1981). In addition, some researchers have claimed to observe that the great apes are able to produce certain kinds of iconic gestures (e.g., Douglas & Moscovice, 2015; Genty & Zuberbühler, 2015; McNeill, 2012; Perlman & Gibbs, 2013; Perlman, Tanner, & King, 2012; Tanner & Byrne, 1996; Russon & Andrews, 2010; but see Call & Tomasello, 2007).

In addition to these arguments in favor of gesture, gesture origins theories are also typically based on two *negative* claims that weigh against vocalizations. The first is that the vocal behavior of extant great apes – and presumably our common ancestor – is highly constrained, restricted to a fixed repertoire of innate, species-typical calls. In contrast to their gestures, it is thought that these calls are produced reflexively in response to specific environmental, emotional triggers, and they are not subject to voluntary control. This limitation suggests that vocalizations, unlike gestures, could not have supported the first stages of language evolution.

The second negative claim is that, unlike gestures, vocalizations lack any substantial potential to ground meaning through *iconicity*, or resemblance between form and meaning. For example, gestures readily afford the iconic representation of actions or spatial relationships between entities, but it is difficult to envisage how vocalizations may afford the same motivated form of representation. Consequently, it is argued that vocalizations are forced to be arbitrary (e.g. Hockett, 1978), giving rise to the symbol-grounding problem (cf. Harnad, 1990). How does the form of a symbol originally come to be connected with a meaning? The critical need to establish a connection between symbol and meaning motivates the theory that iconic (and indexical) gestures are necessary to bootstrap the creation of an arbitrary vocal symbol system (Armstrong &

Wilcox, 2007; Fay, Arbib, & Garrod, 2013; Fay, Lister, Ellison, & Goldin-Meadow, 2014).

However, as the maxim goes: Absence of evidence is not evidence of absence. Indeed, as I present below, there is now substantial evidence to contradict both of these negative claims. Thus, my aim in this paper is to debunk these two "myths" against vocal origins of language. I will show that, in fact, the great apes have considerable ability to voluntarily control their vocalizations, and that they are able to modify their vocalizations, as well as learn entirely new vocal behaviors. I will also show that the vocal modality has rich potential to communicate a diverse array of meanings through iconicity. With these myths against vocal origins debunked, I propose that the seeds of language are rooted in iconicity in both gesture *and* vocalization, going as far back as our last common ancestor with great apes.

Myth 1: Great ape vocal behavior is fixed and reflexive

Whereas the gestural communication of great apes is highly flexible, it is widely believed that their vocal communication is extremely constrained. According to the fixed reflex myth, great ape vocal behavior is characterized as limited to a fixed repertoire of innate, species-typical, reflexively produced calls. Thus it is assumed that great apes lack voluntary control over their vocalizations, and also that they are incapable of vocal learning – unable to learn new vocalizations or even modify old ones. Fitch (2010) observed that this theory appears to have taken root in the behaviorist tradition with Skinner (1957). Skinner described nonhuman vocal behavior as "innate responses" that "comprise reflex systems which are difficult, if not impossible to modify by operant

reinforcement," suggesting further that vocal behavior "below the human level is especially refractory" (p. 463). Such a vocal deficit served as part of the explanation of why the young apes in early cross-fostering experiments did not learn to speak (Furness, 1916; Kellogg & Kellogg, 1933; Hayes, 1951; Witmer, 1909). As Pinker explained (1994, p. 334), the subjects "were at a disadvantage: they were forced to use their vocal apparatus, which was not designed for speech and which they could not voluntarily control." Goodall's (1986) prominent observation of naturally living chimpanzees in the Gombe Reserve has also reinforced this notion. Goodall remarked that "the production of sound in the absence of the appropriate emotional state seems to be an almost impossible task for a chimpanzee" (p. 125). And more recently, Tomasello and Call (2008) summed it up: "There are no reliable reports of major feats of primate vocal learning or invention, or highly flexible vocal use, anywhere in the literature" (pp. 225-226).

Part of the reason that this earlier work came to such strong negative conclusions may have been its theoretical biases towards a categorical and relatively narrow comparison between ape vocal behavior and human speech. In contrast, more recent research has tended to emphasize the need to consider the different components of sound production by the vocal tract (Fitch, 2010; Lameira, 2013; MacLarnon & Hewitt, 1999; Owren, 2011; Pisanki, et al., 2016). Human speech requires fine, coordinated control over three major muscular systems: the respiratory system, including the diaphragm and abdominal muscles; the laryngeal muscles to control the vocal folds; and the oro-mandibular muscles. Apes may exercise varying degrees of control over these different systems, and therefore, it is important to examine each of them and how they are coordinated together.

For example, a critical question that emerges from this componential view is whether great apes are able to flexibly control their larynx to generate sound by vocal fold vibration (i.e., true "vocalization"). Some major theories of speech evolution have emphasized the evolution of fine voluntary control of the larynx as a crucial step that distinguishes human speech from the behavior of great apes (Ackermann, Hage, & Ziegler, 2014; Fitch, 2010). However, in speech, while vocal fold vibration is necessary to produce vowels, many consonants are voiceless (Lameira et al., 2013), and on average, languages have three times as many consonants as vowels (Maddieson, 2011). These considerations highlight the crucial point that the evolution of speech entails fine coordinated control over the respiratory system and supra-laryngeal articulators, in addition to the larynx.

Another consequence of the componential view is that an expanded range of behaviors becomes relevant for study and comparison. Much earlier work in this area stemmed from a research paradigm focused on the prototypically vocal (i.e. with vocal fold vibration), emotional, urgent calls of primates and other mammals (Lameira et al., 2013). Yet, in comparison, humans perform a myriad of actions with their vocal tract, many of which are not obviously communicative or even result in any salient sound. These include multifarious activities like blowing out candles, huffing on eyeglasses, blowing up balloons, blowing bubbles, holding one's breath, gurgling, blowing one's nose, sucking and blowing through straws, and so on. While these behaviors are not vocalizations per se, they nevertheless require dexterous control over components of the vocal tract. They also illustrate the importance of culture in the development of human vocal behavior beyond just the acquisition of spoken language. Therefore, it stands to

reason that an assessment of great ape vocal abilities ought to consider their potential in a cultural context, and look beyond just those vocal behaviors that are clearly for communication. As we shall see, the significance of this point is highlighted by studies of great apes reared in close contact with humans, especially so-called enculturated apes, who develop immersed in human interaction and the surrounding culture.

Along with this theoretical move towards a componential view of vocalizations and speech, a growing number of studies have provided strong evidence to reject the traditional idea that ape vocal behavior is fixed and reflexive. In what follows in this section, I survey the mounting evidence across species and rearing conditions showing that the great apes wield considerable voluntary control over the components of their vocal tract, and also that they are capable of vocal learning.

Voluntary control of vocal behavior

Preliminary evidence that the great apes can exercise voluntary control over their vocal behavior came from Goodall's (1986) field observations of male chimpanzees at the Gombe Reserve (ironically, from the same source as the quote above). Goodall noted that males became unusually silent during patrols around the borders of their territory. This sort of vocal control – perhaps more aptly, inhibition – was also shown in a playback experiment in which wild chimpanzees were played panthoot calls from an unfamiliar male (Wilson et al., 2001). The likelihood that an individual responded vocally to the call was found to depend on the number of allies present, suggesting that the pant hoots were produced strategically. Another study found that wild female chimpanzees were selective in producing pant grunt greetings, tending not to issue a

greeting to an individual if a more dominant individual was present in the vicinity (Laporte & Zuberbühler, 2010).

A study by Schel et al. (2013) aimed to assess whether chimpanzees produce vocalizations intentionally according to the same criteria used for gestural communication. Experimenters placed a moving python model into the path of wild chimpanzees, both when they were travelling alone and when they were with a group. In several ways, the chimpanzees showed evidence that they were intentional in their use of alarm calls to warn others of the snake. They appeared to produce the calls to inform specific individuals, being more likely to vocalize to friends, kin, or higher-ranking individuals. They showed gaze alternation – looking back and forth between their audience and the snake. And they also appeared to produce vocalizations that were goal directed, only ceasing to vocalize when the recipients were safe from the predator.

Experimental studies with captive chimpanzees show that they are able to produce vocalizations and other sounds intentionally to get the attention of human caregivers (Taglialatela et al. 2011; Leavens, Russell, & Hopkins, 2010). Subjects were more likely to produce vocal sounds – a lip sputter and an extended grunt – when they lacked the visual attention of an experimenter, indicating voluntary, thoughtful control over their production. Some recent experiments with captive orangutans leave little doubt that individuals are able to produce voiceless sounds and vocalizations voluntarily. Lameira and colleagues (Lameira et al., 2013; Wich et al., 2009) have documented 10 different cases in which captive orangutans learned to whistle, controlling egressive and ingressive air flow. In most cases, they could whistle on cue, and two of the animals could control their whistle to match the duration and number of whistles of a human model. And

finally, an experiment with a captive, young male orangutan (Rocky) may be the most impressive display of vocal control to date (Lameira, Hardus, Mielke, Wich, & Shumaker, 2016). Rocky produced an idiosyncratic vocalization called a "wookie" by the authors – distinct from any species-specific vocalization – which he used to gain the attention of human caregivers. Trained to play a "do-as-I-do" imitation game, the orangutan produced the wookie in response to the demonstrator, and was even able to modulate his voice frequency to match the high or low frequency of the model. These findings provide strong evidence that orangutans have skillful, voluntary control over their breathing, supralaryngeal articulators, and larynx.

Learning in vocal behavior

Thus the great apes – from orangutans to chimpanzees – possess significant voluntary control over sound production with their vocal tract. In addition, a number of studies show that great apes also have some capacity for vocal learning. They are able to learn to produce vocalizations and voiceless calls that are modified from the species typical calls, or even novel sounds outside of the typical repertoire. As I present below, the literature contains a growing array of examples of vocal learning by great apes.

In wild chimpanzees, the ability to modify species-typical calls is illustrated by studies showing evidence that different populations of chimpanzees have different "dialects" of pant hoots (Crockford et al. 2004; Marshall et al. 1999, Mitani & Gros-Louis, 1998). Individuals within a group tend to produce more similar pant hoots than individuals between different groups, even after controlling for heritability and environmental factors. Call modification was also found in a study of captive

chimpanzees that examined their referential food grunts before and after one group was relocated to live with another group (Watson et al., 2015). Individuals from the new group slowly adjusted their grunt for 'apple' to match the original group, but only after they became socially integrated with members of the original group and strong affiliative relationships were established between them. Critically, while the vocalizations of the newly introduced chimpanzees converged on those of the original group, their preference for apples remained consistent, thus ruling out arousal as a reason for the change.

A more extraordinary case of the capacity to reformulate species-typical food calls comes from the enculturated male bonobo Kanzi, who regularly interacted with humans using a lexigram board, pictures, and (receptively) spoken English (Hopkins & Savage-Rumbaugh, 1991; Taglialatela et al., 2003). Kanzi was found to produce four distinct peep calls that differed from his species-typical peeps in acoustic properties such as duration and pitch contour. Notably, he produced each of the peeps systematically in a particular context, which appeared to correspond to the meanings 'banana', 'grape', 'juice', and 'yes'.Beyond the modulation of species-typical vocalizations, studies also document an ever-growing list of novel vocalizations, voiceless calls, and other vocal behaviors that are outside of the innately specified repertoire. For example, Pika (2014) described four oral sounds produced by chimpanzees of the Ngogo community during grooming interactions. These included: *lip-bobs*, produced by repeatedly and subtly touching the lips to produce a soft round sound; *lip-smacks*, by rapidly and repeatedly touching lips and using the tongue to make a smacking sound; *raspberries*, by pressing air and saliva through the lips to produce a spluttering sound, and *teeth-clacking,* by moving the upper and lower jaws against each other in small movements to make

clicking sounds. Studies have also observed novel voiceless calls used by certain populations of free-ranging orangutans (van Schaik et al., 2006). For example, different groups learned to use a particular variation of a lip sputter during specific phases of an evening nest building procedure. Notably, the call and its particular usage appears to be spread through social transmission.

The catalogue of novel vocal behaviors expands markedly when one considers great apes that were reared in human-intensive environments, especially language-trained and other deeply enculturated animals. Some of this influence by humans appears to come from extensive exposure to speech. For example, Tilda, an ex-entertainment orangutan, appears to have learned to produce two novel calls by imitating temporal qualities of the human speech she heard around her (Lameira et al., 2015). Tilda produced these calls – a voiceless click and vocal "faux-speech" – with a speech-like rhythm of 5 Hz that was faster than other rhythmic calls documented in orangutans. More generally, these calls were unlike any other known orangutan calls, in form, or in context of use, and thus they do not appear to be variations of species-typical calls. And mentioned above, the captive orangutan Rocky produced a novel 'wookie' vocalization, which he was able to voice at different frequencies in imitation of a human model (Lameira et al., 2016).

However, exposure to speech (and vocal imitation games) is the not the only reason that great apes expand their vocal repertoire in human-intensive environments. Another factor is that these environments often include various physical artifacts that scaffold the performance of particular vocal and breathing-related behaviors, e.g. musical instruments, lit candles on a birthday cake, drinks with straws, and so on. Many of the

associated behaviors are not directly communicative or vocal, and they may not even be distinctly audible. But they nevertheless show a capacity to learn new behaviors involving coordinated control over the different components of the vocal tract.

For example, there are several cases in which great apes have learned to control their breathing as part of the performance of behaviors acquired from their experiences with humans and their artifacts. For instance, one report described two individuals – a chimpanzee and an orangutan kept as household pets – that learned to hold their breath while swimming and diving underwater (Bender & Bender, 2013). Both animals could submerge themselves for at least 15 seconds. Breath control is also exhibited in cases in which captive apes have developed an inclination for smoking cigarettes (Witmer, 1909; Kearton, 1925)[1]..Notably, to draw smoke, the smoking ape must combine breath control with the coordinated ability to configure its mouth to form a suitable constriction with the cigarette. And in perhaps an even more impressive example of this type of skill, the bonobo Kanzi is able to blow up a balloon, which he demonstrated during an interview for the Oprah Winfrey show[2] .

The enculturated gorilla Koko offers another remarkable demonstration of oral manipulation and breath control in her "play" with musical wind instruments (Perlman & Clark, 2015; Perlman, Patterson & Cohn, 2012). Koko – the subject of a long-term project to teach her to use symbolic gestures based on American Sign Language – was reared in an intensively human environment from the age of one year (Patterson & Linden, 1981). Video records of Koko as an adult from 37-39 years of age show her

[1] See a smoking orangutan in a Malaysian zoo: https://www.youtube.com/watch?v=B-h_JlpdRJQ

[2] http://www.oprah.com/oprahshow/kanzi-the-talking-ape-video; see 1m50s

blowing notes with a variety of devices, including recorders, harmonicas, and party favor whistles. Video also shows her making sound by blowing across the open end of paper towel tubes, bottles, and even a pen top.

In addition to her sound production with instruments, Koko has learned a repertoire of several behaviors that involve control over her breathing and oro-mandibular articulators (Perlman & Clark, 2015). For example, she blows into her hand, huffs on eyeglasses in a human-like cleaning routine, blows "raspberries" (curling her tongue lengthwise and blowing air through it), blows her nose into a tissue, drinks through straws, and blows out candles on her birthday cake. Koko performs another behavior known as the *blow test* to greet visitors through the mesh of her enclosure. She leans forward toward the visitor and blows gently at their face, apparently as an invitation for them to blow back so that she can smell their breath. In contrast, Koko performs a more aggressive action known as *you blew it* when she is especially agitated at someone. She inhales a great amount of air, and then exhales it forcefully at the subject of her agitation. Koko has also learned behaviors that involve control over her larynx (Perlman & Clark, 2015). On the verbal suggestion of a caregiver, Koko can perform a fake cough (covering her mouth with her hand) – an action that requires her to abruptly constrict her larynx. She also voices variable series of grunts into telephones and long objects that bear a likeness to microphones. Table 1 shows the behaviors of Koko's repertoire and the anatomical control they typically require.

Table 1. Koko's repertoire of behaviors involving breath control. Check marks indicate that the specified articulator plays a prominent role in behavior.

Behavior	Breath	Lips	Tongue	Jaw	Velum	Larynx
Play instrument	✔	✔				
Blow (on hand)	✔	✔				
Cough	✔					✔
"Raspberry"	✔		✔			
Blow nose	✔				✔	
Huff on glasses	✔			✔		
Talk on phone or mic.	✔	✔		✔		✔
"Blow test"	✔	✔				
"You blew it"	✔	✔				
Blow out candles	✔	✔				
Drink through straw	✔	✔				

In all of these different examples – from Koko and Kanzi to wild populations – we see evidence that great apes possess considerable control over their vocal tract, including coordinated control over their breathing, supralaryngeal articulators, and in some cases, their larynx. They are able to modify old vocal behaviors and learn new ones, and even show some capacity for on-the-spot vocal imitation. So given these impressive vocal abilities, especially in cases of enculturated apes, why did longitudinal studies like

Hayes and Hayes with the young chimpanzee Viki find such limited evidence of vocal learning (Hayes, 1951)? (Also see comparable findings in attempts to teach orangutans to speak: Furness, 1916; Laidler, 1980.)

Lameira et al. (2016) pointed out that these cross-fostering ape language projects tended to focus on speech as the benchmark of comparison, rather than the animal's own natural vocal repertoire, preferences, predispositions, and constraints. Consequently, these projects overlooked much of the flexibility in their subjects' vocal behavior. For example, from this more attuned perspective, it is worth noting that according to Hayes (1951), Viki did indeed show evidence of voluntary control over her vocal tract and of some capacity for vocal learning. Initially with training, she learned to vocalize sounds at will, and then she was able to learn some speech sounds by imitating the model of her parents. She was eventually able to articulate four words: 'mama', 'papa', 'cup', and 'up', which combined four or five distinct phonemes[3]. Viki also learned to blow air through her vibrating lips, produce additional sounds transcribed as ''blook'' and ''boo'', and blow spit bubbles. Thus even the notorious case of Viki reveals evidence against the myth that great ape vocal behavior is entirely fixed and reflexive.

Myth 2: Vocalizations cannot be iconic

The second traditional argument against vocal origins of language is based on the myth that, unlike gestures, vocalizations lack any substantial potential for iconic expression. On this idea, gestures are readily used for the iconic representation of spatial relationships, deixis, actions, shapes, and so on (e.g. Cartmill, Beilock, & Goldin-

[3] Cf. the chimpanzee Johnny saying "mama": https://www.youtube.com/watch?v=y4Z0xn4pYSY

Meadow, 2012). But in comparison, vocalizations are extremely limited, only useful for expressing emotion and or for vocal mimicry of animal sounds. As Hockett (1978: 274) put it: "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." In later work, Hockett depicted this notion in a figure with drawings of an elephant and a woodshed in one dimension – each as a line, and rather indistinguishable (1987, p. 11).

The significance of such an iconicity deficit in language evolution is illustrated by Tomasello (2008) in his variation on a classic thought experiment. Imagine that each of two groups of children grow up alone on an island, without any exposure to a language (but are somehow well cared for). The critical variable is that one group (somehow) is only permitted to communicate with vocalizations, while the other group can only communicate with gestures. How would the vocal children compare to the gesturing children in their proclivity to develop a language? Tomasello reasons (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.

It follows from this reasoning that vocalizations must have been "piggybacked" (Tomasello, 2008, p. 330) or "boot-strapped" (Fay et al., 2013) on gestures because of their "vastly greater possibility for iconic productivity in the visual medium" (Armstrong & Wilcox, 2007, p. 123).

This view of vocalization falls in line with the Saussurean principle that spoken words bear an arbitrary relationship between form and meaning (de Saussure, 1959; but see Joseph (2015) on the significant role that iconicity actually played in de Saussure's linguistic theory). This foundational principle was later incorporated into Hocket's (1960) postulation of the fundamental design features of language. Pinker and Bloom illustrated this linguistic principle by explaining, "There is no reason for you to call a dog *dog* rather than cat except for the fact that everyone else is doing it" (1990, p. 718). On this idea, the number of iconic words, such as onomatopoeia and other imitative words is "vanishingly small" (Newmeyer, 1992, p. 758).

Contrary to these dismissive claims, there is abundant evidence showing the great potential for iconicity in vocalization and speech (Clark, 2016; Imai & Kita, 2014; Lockwood & Dingemanse, 2015; Perlman & Cain, 2014). Some of this evidence comes from studies of natural spoken languages, including the prevalence of iconicity in the ideophones of many diverse languages (Dingemanse, 2012), as well as in certain words that tend to have similar forms across languages (Blasi, et al., 2016). And in recent years, a lot of research has focused on a few well-established cases of sound symbolism – especially the so-called "bouba-kiki" effect – to gain deeper understanding of how iconicity functions in word processing and word learning (see Lockwood & Dingemanse, 2015 for review). However, my focus here is to highlight experimental studies showing the rich, semantically diverse potential for iconicity in vocalization. This research is of two types. First, there are a number of studies – often under the umbrella of sound symbolism – showing that different people are consistent with each other in the meanings they derive from the phonemes of novel words. Importantly, people's intuitions for sound

symbolism span a diverse range of meanings and speech sounds. The second line of research examines people's ability to improvise iconic vocalizations to express different meanings, and in turn, whether the improvised vocalizations can be understood by listeners.

Sound symbolism

Numerous studies of sound symbolism demonstrate that people often share consistent intuitions about the meanings of sub-morphemic speech sounds when they are presented in the context of non-sense words (Lockwood & Dingemanse, 2015). For example, in a seminal study by Sapir (1929), participants listened to pairs of nonsense words that contrasted minimally by vowel (e.g. "mil" vs. "mal"), and judged which word referred to the larger version of an arbitrary item (e.g. chair). In general, people judge words with high front vowels (e.g. /i/), which have higher-pitched second formants, to be *small*, and low, back vowels (e.g. /a/), with lower-pitched second formants, to be *large* (Ohala, 1994). Similar intuitions about size-sound symbolism have been found in various experiments and extended to more diverse meanings that may be connected with size, such as *speed* and *quickness*, and attitudes like *affection, intimacy, disdain,* and *dominance* (Bentley & Varon, 1933; Cuskley, 2013; Jespersen, 1933; Newman, 1933; Thompson & Estes, 2011).

In the same year as Sapir's study, Köhler (1929) identified a robust case of shape-sound symbolism, finding that participants overwhelmingly matched the word "takete" with a pointed shape and "baluma" with a rounded shape. Revamped by Ramachandran and Hubbard (2001), Köhler's study eventually gave rise to the many experiments on the

"bouba-kiki" effect. Although the particular mappings that underlie this effect are the subject of ongoing research, the basic account is that voiceless obstruent, palatal consonants alternated with front, closed and unrounded vowels evoke an angular meaning, whereas bilabial, voiced sonorants combined with front, open, and rounded vowels evoke a rounded meaning (Ahlner & Zlatev, 2011). This sort of sound-shape mapping may be flexible, however, as in one study that found people were faster to learn to categorize rounded and pointy-shaped aliens when they corresponded with the labels "crelch" and "foove" (Lupyan & Casasanto, 2014). Variations on the bouba-kiki effect have also been demonstrated with participants across different cultures, including the Himba of Northern Namibia and (Bremner et al. 2013) and the Taiwanese (Chen, Huang, Woods, & Spence, 2016).

Beyond size and shape, experiments have found evidence for various other mappings between speech sounds and different kinds of meanings or stimulus properties. For example, Hirata et al. (2011) found that voiceless consonants were associated with a bright visual stimulus, whereas voiced consonants were associated with a dark one. Moos et al. (2014) found that participants associated higher vowels (with higher second formants) with *redness* in the color spectrum and lower vowels (with lower second formants) with *yellowness*. Another study found that people associate vowel quality and taste (Simner et al., 2010). Participants assigned higher, more back vowels with *sweet*, lower more front vowels with *sour*, and *salty* and *bitter* with vowels in between.

Other studies have taken a more exploratory approach. Greenberg and Jenkins (1966) asked participants to rate English consonants and vowels along several semantic dimensions. Different groupings of consonants were consistently distinguished according

to properties that included *abrupt-continuous, liquid-solid, tight-loose, delicate-rugged, angular-rounded, active-passive, good-bad,* and *inhibited-free.* Vowels, corresponding with tongue position, were distinguished along different dimensions: *high-low*, *sharp-dull*, *narrow-wide, thick-thin, oblong-round, large-small,* and *falling-rising*. Similar studies have been conducted in marketing research on the sound symbolic associations of product brand names. For example, in one study, participants completed a questionnaire with questions like "Which brand of ketchup seems thicker? Nidax or Nodax? (Klink, 2000, p. 11; also see Kelly, Leben, & Cohen, 2004). The results showed that brand names with front vowels were judged as *smaller, lighter (vs. darker), milder, thinner, softer, faster, colder, more bitter, more feminine, 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.*

Iconic vocalizations

Studies of sound symbolism show that people are able to derive a diversity of meanings from a wide span of speech sounds. People also share intuitions of how to improvise iconic vocalizations to express various meanings. Evidence of this comes primarily from recent experiments using charades-style communication games to examine the potential for people to improvise iconic vocalizations and vocal pantomimes. For example, one set of studies aimed to directly compare the use of non-linguistic vocalization and gesture for iconic representation (Fay, Arbib, & Garrod, 2013; Fay, Lister, Ellison, & Goldin-Meadow, 2014). Participants communicated 18 different items

to a partner who attempted to guess the referent. The items included emotions (e.g. disgust, tired), actions (e.g. throwing, chasing) and objects (e.g., rock, fruit). The main conclusion of the studies was that participants performed better with gestures than vocalizations. However, players showed success with vocalizations, particularly for emotions and actions, which were identified at accuracy levels that well exceeded chance.

Another set of experiments has examined more specifically what kinds of sounds people produce when they vocalize various meanings (Perlman & Cain, 2014; Perlman, Dale, & Lupyan, 2015; Perlman, Paul, & Lupyan, 2015). Do different people consistently produce vocalizations with similar acoustic properties for particular meanings? By comparable logic to the shared intuitions people demonstrate in interpreting sound symbolism, this would suggest that their vocalizations are iconic, conveying a sense of meaning directly through their form. In these experiments, pairs of participants took turns improvising vocalizations to communicate different meanings to a guessing partner. The rules of the game emphasized that participants were not permitted to use words or to make vocalizations that sounded like particular words.

For example, in an initial exploratory study, participants communicated 60 different meanings to a partner (Perlman & Cain, 2014). Each player held a set of 30 cards, which contained mixed within it 15 words and their antonyms. These included items like *alive, dead, dull, sharp, now, later, hard, soft, few, many, fast, slow, gad, good, bright,* and *dark*. The acoustic properties of each vocalization were measured, including its mean pitch, pitch change, pitch range, intensity, duration, harmonics to noise ratio, and repetition rate. The results showed that participants shared a sense of how to distinguish 20 of the 30 antonymic word pairs according to similar acoustic properties

(26 of 30 without correcting for repeated tests). For example, *rough* compared to *smooth* was communicated with aperiodic sounds, *small* compared to *large* with quiet, high-pitched sounds, and *fast* compared to *slow* with loud high-pitched, quickly repeated sounds. Notably, participants used each of the seven acoustic properties to express at least one meaning or another. Table 3 presents examples from different vocal charades experiments of antonymic meanings and the acoustic properties used to distinguish them.

In another study, participants played a 10-round iterated version of vocal charades, this time communicating meanings from a reduced set of nine opposite pairs. These included *attractive / ugly, bad / good, big / small, down / up, far / near, fast / slow, few / many, long / short,* and *rough* / *smooth*. With few exceptions, each meaning was expressed with characteristic acoustic properties that distinguished it from each other meaning. In subsequent playback experiments, the vocalizations were played for naïve listeners on Amazon Mechanical Turk, who were quite successful at guessing their meanings. Their accuracy was at least 20% compared to a chance rate of 10% for 15 of the 18 different meanings. Moreover, the likelihood that listeners guessed a particular meaning for a vocalization could be predicted based on how the vocalization compared to an iconic "template" for that meaning. The template was derived from the specific properties that the charades players reliably used to distinguish the meaning from its antonym. The more a vocalization exhibited extreme values of these iconic properties for a given meaning, the more likely naïve listeners were to guess that meaning. In addition, not only could participants create iconic vocalizations, they were also able to conventionalize these into more word-like symbols through repeated interactions. Over the course of ten rounds of vocal charades, what began as relatively variable improvised

vocalizations became shorter in duration and more stable in form. At the same time, the vocalizations were also understood more rapidly and accurately.

An even wider array of meanings were tested in a third vocal charades study, in which participants were invited to compete in the "Vocal Iconicity Challenge!" (Perlman & Lupyan, 2015)[4]. Eleven competing teams submitted recorded vocalizations for 30 different meanings, which included 8 actions (e.g. cook, gather, hide), 12 things (e.g. deer, water, knife), and 10 properties (e.g. sharp, big, many). These vocalizations were played to naïve listeners on Amazon Mechanical Turk who guessed their meanings. The team whose vocalizations were guessed most accurately was crowned *Vocal Iconicity Champion* and winner of the $1000 *Saussure Prize.* The results showed that listeners were more accurate than chance at guessing vocalizations for each of the 30 different meanings. With a chance guessing rate at about 10%, accuracy for the vocalizations of six of the eleven teams was over 40%, with a rate of 57% for the winning team.

Table 2. Examples of typical iconic vocalizations produced in vocal charades studies.

Word pair	Semantic domain	Acoustic properties	Study
Alive vs. Dead	Animacy	More harmonic, more loud	P&C2014
Smooth vs. Rough	Texture	More harmonic, higher pitch, lower intensity	P&C2014, PD&L2015
Down vs. Up	Space	Falling pitch, lower pitch, longer duration	P&C2014, PD&L2015
Rock vs. Fruit	Objects	Shorter in duration,	P&L2015

[4] see http://sapir.psych.wisc.edu/vocal-iconicity-challenge/

Sharp vs. Dull	Shape	Higher pitch, shorter duration	P&C2014, P&L2015
Good vs. Bad	Appraisal	More harmonic, larger pitch range, higher pitch	P&C2014, PD&L2015
This vs. That	Deixis	Shorter duration, less decrease in pitch	P&L2015
Man vs. Woman	Gender	Less harmonic, lower pitch, less decrease in pitch	P&C2014; P&L2015
Many vs. One / Few	Number	Faster repetition of sounds, longer duration	P&C2014; P&L2015
Cut vs. Pound	Action	Lower intensity, higher pitch	P&L2015

Note: Acoustic properties distinguish the first word from the second. P&C2014 = Perlman & Cain (2014), PD&L2015 = Perlman, Dale, & Lupyan (2015) and P&L2015 = Perlman & Lupyan (2015).

One limitation of these vocal charades studies is that they were conducted with English-speaking participants. An important direction for future work is to examine the ability of people to improvise iconic vocalizations that communicate successfully with listeners from disparate cultures and linguistic backgrounds. For example, one case in which the ability to produce iconic vocalizations appears to be highly robust across cultures – at least in some respects – is the domain of magnitude. Two groups of children living in China played a game in which they were asked to improvise vocalizations to communicate between items contrasting along different dimensions of magnitude, for

example, a *long* versus *short* string and a *big* versus *small* ball (Perlman et al., 2015). One group consisted of hearing children 9 to 12 years of age, and the second group consisted of deaf children from 7 to 20 years of age.

The results showed that both groups – hearing and deaf participants alike – consistently produced vocalizations that were longer in duration and louder for the greater-magnitude items. Both groups were also somewhat more nuanced in their vocalizations, for example, especially extending vocal duration for the *long* string in particular. However, only the hearing children consistently used the pitch (i.e. fundamental frequency) of their voice, and perhaps surprisingly, their inclination was to use higher pitch for the larger-magnitude items – in contrast to previous results with English-speaking participants (e.g. Perlman & Cain, 2014). Subsequent work played the vocalizations from both groups of Chinese participants to American listeners, who were able to guess the intended referent of the vocalizations from both groups at greater-than-chance levels (Perlman, Paul, & Lupyan, under review). Mirroring the production results, listeners relied most consistently on duration and loudness in selecting the item, but were inconsistent in using pitch as a cue. These findings suggest that some iconic mappings in vocalizations – such as between duration and loudness, and magnitude – may be highly robust across culture and experience, whereas other mappings – for example, between pitch and magnitude – may be more culturally variable and experience dependent.

Altogether, these studies on sound symbolism and the ability to improvise iconic vocalizations demonstrate the rich potential for iconicity in the vocal modality. Thus, contrary to the myth, vocalizations can be iconic. This presents a fatal challenge to Tomasello's (2008) argument against vocal origins of language that vocalizations do not

enable the creation of "already meaningful communicative acts." People can both understand iconic vocalizations and novel words, and they create new iconic vocalizations for meanings spanning a wide range of conceptual domains.

Conclusion: Language is iconic and multimodal to the core

In this paper, I have aimed to debunk two common arguments raised against proposals for vocal origins of language. The first relates to the vocal abilities of the great apes. Rather than being fixed and reflexive, research increasingly shows that apes have considerable voluntary control over their vocal behavior, including their breathing, larynx, and supralaryngeal articulators. They are able to learn new vocal behaviors, and even show some rudimentary ability to spontaneously imitate elements like the pitch, duration, and number of repeated calls. Second, contrary to the idea that vocalizations cannot be iconic, an abundance of research demonstrates that the vocal modality affords rich potential for iconicity. People can understand sound symbolism, and they can produce iconic vocalizations for a diverse range of meanings. Thus, two of the primary arguments against vocalizations – and in favor of gesture-first theories – are not tenable.

Moreover, although there is not space here for detailed discussion, there is also growing evidence that modern spoken languages are actually quite iconic (Dingemanse et al., 2015; Imai & Kita, 2014; Perlman & Cain, 2014; Perniss & Vigliocco, 2014). This includes iconicity in basic grammatical patterns, such as the parallel between the order of linguistic elements and their order in physical experience (as in telling a story), and the correspondence between increased morphological complexity and increased semantic complexity (Haiman, 1980; Lewis & Frank, 2016). Iconicity is prevalent in the lexicons

of many languages across the world, which have large inventories of ideophones and onomatopoeic words that are used to express a variety of sensory-rich meanings (Dingemanse, 2012). And a recent statistical analysis reported that a considerable proportion of 100 basic vocabulary items across a large diverse sampling of languages carries strong associations with particular speech sounds, suggesting iconicity (Blasi et al., 2016). Even English, which is notoriously lacking in iconic words (e.g. Perniss & Vigliocco, 2014), shows evidence that iconicity plays a significant role in the structure of vocabulary, especially in words learned earliest by children (Perry, Perlman, & Lupyan, 2015) and in words with highly sensory meanings (Winter, Perlman, Perry, & Lupyan, this issue). These findings showing that iconicity is pervasive across spoken languages suggest that it at least plays an ongoing role in their historical development, if not their original inception.

There has been a tendency in previous scholarship on language evolution to pit gestures against vocalizations, with an emphasis on determining which came first. For instance, Fay et al. (2014) highlighted their comparison of modalities with the article title "Gesture beats vocalization hands down." Yet, while I have set out here to dispose of the main arguments against vocal origins, this does not detract from the positive arguments in favor of gestures playing a role in the origins of language. The gesturing of great apes is impressively flexible, and the iconic potential of gestures is well attested (e.g. Arbib, Liebal, & Pika, 2008; Cartmill, Beilock, & Goldin-Meadow, 2012). Thus, the current weight of evidence supports the hypothesis that language began through iconic communication coordinated across both the vocal and gestural modalities.

Notably, this is a different claim than some previous proposals for multimodal origins of language. Some gesture scholars have argued for a multimodal evolution of language in the context of a theory of human communication in which gestures mainly incorporate the imagistic or iconic aspects of expression, while vocalizations serve mainly for the conventionalized, linguistic channel (e.g. Kendon, 2014; McNeill 2012). However, studies of speech prosody (Shintel, Nusbaum, & Okrent, 2006; Perlman, Clark, & Johansson Falck, 2015), the use of ideophones (Dingemanse, 2013), and the use of quotation (Blackwell et al., 2015; Clark & Gerrig, 1990) illustrate how iconicity actively permeates the vocal modality, along with gestures. Clark (2016) argued that depiction is a critical mode of communication, and his examples illustrate the richness of depiction across the modalities.

In addition, many researchers of ape communication are also coming to advocate for a more deeply multimodal evolution of language (Taglialatela et al., 2011; Liebal, Waller, Burrows, & Slocombe, 2013). For example, Leavens (2003) suggested that, "Because visual and vocal communication seem to be functionally linked in extant apes, language may have been multimodal from its inception" (p. 233). In this respect, Koko's repertoire of learned vocal and breathing related behaviors may be a revealing example of the capacity of great apes to produce novel behaviors involving multiple actions that are coordinated across modalities. Of hundreds of video-recorded instances, Koko performed the vast majority within a larger behavioral complex that combined vocal and oral articulatory movements with various manual gestures and praxic actions (Perlman & Clark, 2015). For example, she combined blowing (bilabial frication) with various

manual gestures, including bringing a single open palm to her mouth, bringing both hands perpendicular to her mouth, and bringing both hands over her mouth.

Considering the evidence, I propose that language is rooted in iconicity in both gesture and vocalization, originating from rudimentary capacities present in our last common ancestor with great apes. Although it is debated, there is increasing evidence that great apes are able to produce iconic manual gestures and pantomimes (e.g. Douglas & Moscovice, 2015; Genty & Zuberbühler, 2016; Perlman & Gibbs, 2013; Perlman et al., 2012; Russon & Andrews, 2010). And there is some suggestion that apes may be able to innovate rudimentary sorts of iconic, or at least "motivated" gestures articulated with the vocal tract. For example, with the *breath test*, Koko performs gentle breaths in a friendly breathing ritual compared to a single forceful breath in the reprimanding action of *you blew it*. The contrasting forms of these communicative behaviors reflect the contrasting emotion of their meaning. A potential example of this in the wild comes from a population of orangutans in which members were found to produce kiss squeaks while holding a stripped leaf to their mouth (Hardus, et al., 2009). This modification, which functioned to decrease the fundamental pitch of the sound, tended to be produced by smaller individuals in high distress, suggesting that it was produced to sound bigger and ward off predators.

Thus the great apes appear to possess a rudimentary capacity to innovate new communicative signals with the vocal tract, as well as with gesture. This initial potential to ground an open-ended system of communication arguably gave rise, over millions of years of evolution, to the deeply iconic, multimodal nature of human communication

today. According to this theory, language is multimodal and iconic to its evolutionary core.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37,* 529-604.

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

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology, 49,* 1053-1076.

Armstrong, D. F., & Wilcox, S. E. (2007). *The Gestural Origin of Language*. Oxford ; New York: Oxford University Press.

Bates, E. (1976). Language and context: The acquisition of pragmatics. New York: Academic Press.

Bender, R., & Bender, N. (2013). Brief communication: Swimming and diving behavior in apes (*Pan troglodytes* and *Pongo pygmaeus*): First documented report: Swimming and Diving behavior in Apes. *American Journal of Physical Anthropology*, *152*(1), 156–162. http://doi.org/10.1002/ajpa.22338

Bentley, M. & Varon, E. J. (1933). An accessory study of phonetic symbolism. *American Journal of Psychology, 45,* 76-86.

Blackwell, N.L., Perlman, M., & Fox Tree, J.E. (2015). Quotation as a multimodal construction. *Journal of Pragmatics, 81,* 1-7.

Blasi, D. E., Wichmann, S., Hammarstrom, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*.

Bonvillian, J. D., Orlansky, M. D., & Novack, L. L. (1983). Developmental milestones: Sign language acquisition and motor development. *Child Development, 54,* 1435-1445.

Call, J. & Tomasello, M. (Eds.) (2007) *The gestural communication of apes and monkeys*. London: Lawrence Erlbaum Associates, Publishers.

Cartmill, E. A., Beilock, S. & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. http://doi.org/10.1037/rev0000026

Clark, H. H., & Gerrig, R. J. (1990). Quotations as Demonstrations. *Language*, *66*(4), 764. http://doi.org/10.2307/414729

Clark, N., Perlman, M., & Johansson Falck, M. (2014). Iconic Pitch Expresses Vertical Space. In *Language and the Creative Mind* (pp. 393–410). Stanford: SCLI Publications.

Corballis, M. C. (2003). *From hand to mouth: the origins of language*. Princeton, N.J.; Woodstock: Princeton University Press.

Crockford C, Herbinger I, Vigilant L, Boesch C (2004) Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology, 110*, 221–243.

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

Dingemanse, M. (2012). Advances in the Cross-Linguistic Study of Ideophones: Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass*, *6*(10), 654–672. http://doi.org/10.1002/lnc3.361

Dingemanse, M. (2013). Ideophones and gesture in everyday speech. *Gesture, 13,* 146-165.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19,* 603-615.

Douglas, P. H. & Moscovice, L. R. (2015). Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Scientific Reports, 5,* 13999.

Fay, N., Arbib, M., & Garrod, S. (2013). How to Bootstrap a Human Communication System. *Cognitive Science*, *37*(7), 1356–1367.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in Psychology*, *5,* 1-12.

Feyereisen, P. & de Lannoy, J. D. (1991). *Gestures and speech: Psychological investigations.* Cambridge University Press.

Fitch, W.T. (2010). *The evolution of language.* Cambridge University Press, Cambridge

Fouts, R. S. & Tukel, S. (1997). *Next of kin: My conversations with chimpanzees.* William Morrow.

Furness W.H. (1916) Observations on the mentality of chimpanzees and orang-utans. *Proceedings of the American Philosophical Society, 55*, 281–290.

Genty, E. & Zuberbühler, K. (2015). Iconic gesturing in bonobos. *Communicative & Integrative Biology, 8,* e992742.

Goldin-Meadow, S. (2016). What the hands can tell us about language emergence. *Psychonomic Bulletin and Review*, doi:10.3758/s13423-016-1074-x.

Goodall, J. (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge.

Graham, K.E., Furuichi, T. & Byrne, R.W. The gestural repertoire of the wild bonobo (*Pan paniscus)*: a mutually understood communication system. *Animal Cognition*. doi:10.1007/s10071-016-1035-9

Greenberg, J. H. & Jenkins, J. J. (1966). Studies in the psychological correlates of the sound system of American English. *Word, 22,* 207-242.

Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language, 56,* 515-540.

Hardus, M. E., Lameira, A. R., Van Schaik, C. P. & Wich, S. A. (2009). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society B*, *276,* 3689-3694.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346.

Hayes, C. (1951). *The ape in our house.* Harper, New York.

Hewes, G. W. (1973). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, *14*(1/2), 5–24.

Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner's speeded classification task. *Perceptual & Motor Skills, 113,* 929-940.

Hockett, C. F. (1978). In Search of Jove's Brow. *American Speech*, *53*, 243–313.

Hockett, C. F. (1987). *Refurbishing our foundations*. Amsterdam: John Benjamins.

Hopkins, W. D. & Savage-Rumbaugh, S. (1991). Vocal communication as a function of differential rearing experiences in Pan paniscus: a preliminary report. *International Journal of Primatology, 12*, 559–583.

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

Iverson, J. M. & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396,* 228-229.

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

Joseph, J. E. (2015). Iconicity in Saussure's linguistic work, and why it does not contradict the arbitrariness of the sign. *Historiographia Linguistica, 42,* 85-105.

Kearton, C. (1925). *My friend Toto: The adventures of a chimpanzee, and the story of his journey from the Congo to London.* London: Arrowsmith.

Kellogg, W. N. & Kellogg, L. A. (1933). The ape and the child: a study of environmental influence upon early behavior. Whittelsey House, Oxford.

Kelly, B. F., Leben, W. & Cohen, R. (2003). The meanings of consonants. *Proceedings of the 29th Berkeley Linguistics Society*. (pp. 245-253).

Kendon, A. (1991). Some considerations for a theory of language origins. *Man,* (N.S.) 26, 602-619.

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

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters, 11,* 5-20.

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

Laidler, K. (1980). *The talking ape*. Stein and Day, New York

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., Shumaker, R. W. & Wich, S. A. (2013). Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America, 134,* 2326.

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. *PLoS ONE, 10,* e116136.

Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports, 6,* 30315.

Lameira, A. R., Maddieson, I., & Zuberbühler, K. Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences.*

Laporte, M. N. C. & Zuberbühler, K. (2010). Vocal greeting behavior in wild chimpanzee females. *Animal Behaviour, 80*, 467–473.

Leavens, D. A. (2003) Integration of visual and vocal communication: evidence for Miocene origins. *Behavioral and Brain Sciences, 26*, 232–233.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition, 13*, 33–40.

Levinson S. C. & Holler J (2014) The origin of human multi-modal communication. Philosophical Transactions Royal Society B 369:20130302.

Lewis, M. & Frank, M. (2016). The length of words reflects their conceptual complexity. *Cognition, 153,* 182-195.

Liebal, K., Waller, B., Burrows, A. & Slocombe, K.E. (2013). Primate communication: A multimodal approach. Cambridge University Press.

Lockwood, G. & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology, 6: 1246*.

Lupyan, G. & Casasanto, D. (2014). Meaningless words promote meaningful categorization. *Language and Cognition,* 1-27.

MacLarnon, A. M. & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology, 109,* 341-363.

Maddieson, I. (2011). Consonant-vowel ratio. In *The World Atlas of Language Stuctures Online.* Max Planck Digital Library.

Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the

structure of vocalizations in chimpanzees? *Animal Behaviour, 58*, 825–830

McNeill D (2012). How language began: Gesture and speech in human evolution. Cambridge: Cambridge University Press.

Miles, L. (1990). The cognitive foundations for reference in a signing orangutan. In Parker, S. T. & Gibson, K. R. (Eds.) *"Language" and intelligence in monkeys and apes* (pp. 511-539). Cambridge: Cambridge University Press.

Mitani, J. C. & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: tests of three hypotheses. *Behaviour, 135* 1041–1064

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder. *i-Perception, 5,* 132-142.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology, 45,* 53-75.

Ohala, J. 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.

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. American Journal of Primatology, 73, 530–544

Patterson, F. G. & Linden, E. (1981). *The education of Koko.* New York: Holt, Rinhart & Winston.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350.

Perlman, M. & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition, 18,* 1165-1179.

Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic Prosody in Story Reading. *Cognitive Science*, *39*(6), 1348–1368.

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

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

Perlman, M. & Lupyan, G. (2015). The vocal iconicity challenge! Paper presented at Protolang 4 conference, Rome, Italy.

Perlman, M., Patterson, F. G. & Cohn R. H. (2012). The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biolinguistics, 6,* 433-444.

Perlman, M., Paul, J.Z., & Lupyan, G. (2015). Congenitally deaf children produce iconic vocalizations to communicate magnitude. In Noelle, D.C., Dale, R., Warlaumont, A.S., Yoshimi, J., Matlock, T., Jennings, C.D., & Maglio, P. P. (Eds.) *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1853-1856). Austin, TX Cognitive Science Society.

Perlman, M., Tanner, J. E., & King, B. J. (2012). 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). John Benjamins Publishing Company.

Perniss & 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

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147.

Pika, S. (2014). Chimpanzee grooming gestures and sounds: what might they tell us about how language evolved. In The Social Origins of Language, pp.128-140

Pinker, S. (1994). The language instinct. William Morrow & Co., New York.

Pisanki, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences, 20,* 304-318.

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

Russon, A. & Andrews, K. (2010). Orangutan pantomime: Elaborating the message. *Biology Letters.*

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proceediings of the National Aacademy of Sciences, 102,* 2661-2665.

Sapir (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12,* 225-239.

Schel, A. M., Townsend, S. W., Machanda, Z., ZuberbÃ¼hler, K., & Slocombe, K. E. (2013). Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PLoS ONE*, *8*(10), e76674.

Senghas, A., Kita, S., & zuryek, A. (2004). Children creating core properties of language: evidence from an emergining sign language in Nicaragua. *Science, 305,* 1779-1782.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*(2), 167–177.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception, 39,* 553-569.

Skinner, B. F. (1957). Verbal behavior. Appleton-Century-Crofts, New York.

Taglialatela, J. P., Savage-Rumbaugh, S., & Baker, L. A. (2003). Vocal production by a language-competent *Pan paniscus. International Journal of Primatology, 24,* 1-17.

Taglialatela J. P., Russell J. L., Schaeffer J. A. , Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. PLoS One 6:e18852.

Tanner, J. E. & Byrne, R. W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37,* 162-173.

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

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

Van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I.,

Suzuki, A., Utami, S. S. & Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science, 3,* 102-105.

Watson, S.K. et al. (2015) Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*, 495–499.

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates, 50,* 56-64.

Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2001). Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Animal Behaviour, 61,* 1203-1216.

Witmer, L. (1909). A monkey with a mind. *The Psychological Clinic III,* 179-205.