



# The Structural Effects of Modality on the Rise of Symbolic Language: A Rebuttal of Evolutionary Accounts and a Laboratory Demonstration

Victor J. Boucher<sup>1\*</sup>, Annie C. Gilbert<sup>2,3</sup> and Antonin Rossier-Bisaillon<sup>1</sup>

<sup>1</sup> Laboratoire de Sciences Phonétiques, Département de Linguistique, Université de Montréal, Montreal, QC, Canada,

<sup>2</sup> School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada, <sup>3</sup> Centre for Research on Brain, Language and Music, Montreal, QC, Canada

## OPEN ACCESS

### Edited by:

Jay Rueckl,  
University of Connecticut,  
United States

### Reviewed by:

Lorenzo Magrassi,  
University of Pavia, Italy  
Alexander Bergs,  
Osnabrück University, Germany

### \*Correspondence:

Victor J. Boucher  
victor.boucher@umontreal.ca

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 25 August 2017

**Accepted:** 05 November 2018

**Published:** 28 November 2018

### Citation:

Boucher VJ, Gilbert AC and  
Rossier-Bisaillon A (2018) The  
Structural Effects of Modality on  
the Rise of Symbolic Language:  
A Rebuttal of Evolutionary Accounts  
and a Laboratory Demonstration.  
*Front. Psychol.* 9:2300.  
doi: 10.3389/fpsyg.2018.02300

Why does symbolic communication in humans develop primarily in an oral medium, and how do theories of language origin explain this? Non-human primates, despite their ability to learn and use symbolic signs, do not develop symbols as in oral language. This partly owes to the lack of a direct cortico-motoneuron control of vocalizations in these species compared to humans. Yet such modality-related factors that can impinge on the rise of symbolic language are interpreted differently in two types of evolutionary storylines. (1) Some theories posit that symbolic language originated in a gestural modality, as in “sign languages.” However, this overlooks work on emerging sign and spoken languages showing that gestures and speech shape signs differently. (2) In modality-dependent theories, some emphasize the role of iconic sounds, though these lack the efficiency of arbitrary symbols. Other theorists suggest that ontogenesis serves to identify human-specific mechanisms underlying an evolutionary shift from pitch varying to orally modulated vocalizations (babble). This shift creates numerous oral features that can support efficient symbolic associations. We illustrate this principle using a sound-picture association task with 40 learners who hear words in an unfamiliar language (Mandarin) with and without a filtering of oral features. Symbolic associations arise more rapidly and accurately for sounds containing oral features compared to sounds bearing only pitch features, an effect also reported in experiments with infants. The results imply that, beyond a competence to learn and use symbols, the rise of symbolic language rests on the types of signs that a modality of expression affords.

**Keywords:** language evolution, symbolic communication, neurophysiology of speech, language development, comparative physiology

## INTRODUCTION

There is a vast literature on the origin of spoken language, much of which offers diverging viewpoints with few areas of consensus. For instance, there is no agreement in this literature on how to define “language” (Christiansen and Kirby, 2003). On the other hand, it is widely accepted that a fundamental feature of language is its symbolic function and that, aside from humans, no other species have developed systems of signs such as those that appear in spoken language. Indeed, for

some, humans are the *symbolic species* (Deacon, 1997). However, claiming the human specificity of symbolic communication rests on how one defines symbols, and the processes by which they evolved. The present paper aims at clarifying these processes within evolutionary theories while offering a demonstration of how the ability to articulate sounds presents an essential factor in the rise of symbolic language.

For some readers, this ability relating to a specific modality of expression may seem to be an obvious factor in the rise of symbolic communication. Yet general definitions of symbols often overlook processes of expression and how they contribute to the formation of signs. In fact, many evolutionists refer to Saussure (1916/1966) and Peirce (1998) and define “symbols” principally as arbitrary associations between signals and concepts of objects or events (i.e., “referents”). Authors also recognize that symbolic associations can operate from memory, when designated referents are not in the context of communication, a feature that Hockett (1960) called “displacement.” These criteria are useful in distinguishing symbols from signs that operate as *icons* or *indices*. The latter involve a *non-arbitrary* resemblance or physical connections to referents, whereas nothing in the attributes of symbols provides a clue as to their interpretation (Deacon, 2012, p. 14; Stjernfelt, 2012). However, definitions from Saussure and Pierce which focus only on arbitrary association can lead to a conceptualization of symbols as mental constructs, unrelated to modalities of expression. Indeed, Saussure saw language as reflecting a separate mental capacity or “faculty” that could generate symbols in *any* modality such as speech or gestures (Saussure, 1916/1966, pp. 10–11). Such ideas have had a lasting influence, especially on linguistic theory, where language is seen to reflect a mental competence that has little to do with modalities of performance (Chomsky, 1965; Hauser et al., 2002). But if this is the case, then why is it that symbolic language develops primarily in a vocal medium?

In focusing on this question, the following discussion draws attention to a body of work in primatology which has failed to uncover a distinct mental ability that could account for symbolic language in humans. On the other hand, we outline that humans are the only primates that possess a cortical control of vocal signals so that, overall, the data undermine the belief that symbolic language arose from an amodal mental competence. A review of this belief that underlies popular theories of the origin of spoken language serves as background to a demonstration of an opposing modality-dependent principle where symbolic language is seen as relating to an ability to articulate sounds. Such a demonstration reflects the approach of a group of studies where evolutionary scenarios are submitted to critical laboratory experiments and computer simulations (as in Gasser, 2004; Monaghan and Christiansen, 2006; Oudeyer and Kaplan, 2006; Williams et al., 2008; Monaghan et al., 2011, 2012, 2014).

## Cognitive Skills as Insufficient Factors in the Rise of Symbolic Communication

In reviewing hypotheses of the origin of spoken language, it is important to acknowledge that several cognitive abilities and neural processes which were thought to underlie symbolic

communication in humans have since been observed in other primates. In particular, it has been established that, with training, apes can learn vast sets of visual symbols and can combine these productively (e.g., Savage-Rumbaugh et al., 1993; Savage-Rumbaugh and Lewin, 1996; Lyn and Savage-Rumbaugh, 2000; Savage-Rumbaugh and Fields, 2000). Follow-up studies have documented that chimpanzees and bonobos raised in symbol-rich environments can develop a vocabulary and utterance complexity similar to those of 3 year-old children (Lieberman, 1984, chapter 10; Gardner and Gardner, 1985, 1969; Pedersen, 2012). There are also reported cases where chimpanzees acquired elements of *American Sign Language* (ASL) only by communicating with other ASL-trained chimpanzees (Gardner and Gardner, 1985). Moreover, brain-imaging research indicates that associative memory in symbol learning involves similar neurological structures in apes, monkeys, and humans (e.g., Squire and Zola, 1996; Eichenbaum and Cohen, 2001; Wirth et al., 2003).

Other symbol-related abilities extend to non-human primates despite continuing claims to the contrary. Of note, the capacity to create hierarchical or embedded combinations of signs – said to reflect a process of “recursion” – was held to be uniquely human (Hauser et al., 2002; Fitch et al., 2005; cf. Hauser, 2016). Some also maintained that a related ability to combine symbols based on conceptual relations, a property termed “Merge,” was distinctly human (Chomsky, 1972, 2012; Bolhuis et al., 2012; cf. Lieberman, 2016). However, Perruchet and Rey (2005) demonstrated that chimpanzees can learn to generate embedded sequences of given symbols (and see Perruchet et al., 2004, 2006; Sonnweber et al., 2015). Additionally, research has shown that monkeys can distinguish acoustic cues in speech (as discussed by Belin, 2006), and manifest a “statistical learning” of speech sounds (Hauser et al., 2001). Several reports have further shown that non-human primates can process combinations of symbols based on conceptual relations (contra, e.g., Jackendoff, 2002; Pinker and Jackendoff, 2005; Clark, 2012; Hauser et al., 2014; Arbib, 2015). Thus, seminal work by Savage-Rumbaugh et al. (1986, 2001; Savage-Rumbaugh and Rumbaugh, 1978) revealed that training chimpanzees on paired symbols designating items and actions (“drink” and “liquids” vs. “give” and “solid foods”) facilitated the learning of combinations of novel signs. In other words, the individuals more easily acquired pairs where action symbols correctly matched signs for types of foods, which implies a processing of signs in terms of their conceptual relations (Deacon, 1997, p. 86). More recently, Livingstone et al. (2014) trained rhesus monkeys on symbols representing distinct numbers of drops of liquid (implying a coding of magnitude). On tests involving combinations of these learned symbols, the individuals not only showed a capacity to process the *relative* values of signs within a context, but also transferred these subjective valuations to new symbols, suggesting a capacity to process combined signs in terms of novel relations. Livingstone et al. (2010, 2014) also outlined that value coding of signs involves similar neural processes in humans and monkeys implicating dopamine neurons and interactions between the midbrain, the orbitofrontal cortex, and nucleus accumbens (for a critical review of other findings of this type, see Núñez, 2017).

Finally, it was also believed that only humans have the ability to *imitate*, while apes *emulate* behaviors (e.g., Tomasello, 1990, 1996; Heyes, 1993; Fitch, 2010). Emulation has been characterized as entailing a learning of the effects of actions, rather than a copying of bodily motions. A limited capacity to imitate was thought to hinder the cultural transmission of communicative signs and tool use. Even so, some studies have shown that non-human primates can mimic the actions of their conspecifics and also learn to produce their calls and symbols (Sutton et al., 1973; Gardner and Gardner, 1985, 1969; Pedersen, 2012). Recent research has made it clear that apes can selectively apply a range of social learning processes. This includes *deferred imitation* as well as the ability to follow eye-gaze and direct attention by eye-gaze and pointing (for a review, see Whiten et al., 2009). Other related claims to the effect that only humans have “shared intentionality” and an advanced “theory of mind” (e.g., Pinker and Jackendoff, 2005) have been questioned in studies of apes reared by humans (Bullock et al., 2008; Leavens et al., 2009; see also contra Penn and Povinelli, 2007; Call and Tomasello, 2008).

In short, research in the last decades has revealed that, contrary to held assumptions, non-human primates possess mental abilities that serve to learn and process symbols. But the fact remains that monkeys and apes in the wild do not develop repertoires of symbolic signs of the type used in spoken language (e.g., Pika et al., 2005). To illustrate the kinds of signs that arise in non-human primates, one can consider the often cited case of vervet monkeys who use distinct signals to communicate the presence of different predators (Seyfarth et al., 1980; Cheney and Seyfarth, 1990; Price et al., 2015; for similar referent-specific signals in apes, see Crockford and Boesch, 2003). It has been argued that these signs are symbols based on their seeming arbitrariness (Price et al., 2015), though some critics reject this interpretation (Hauser et al., 2002; Deacon, 2012). In fact, vocal signs of apes and monkeys appear to be largely indexical in that they reflect reactions to referents in the signaling context (see also Crockford and Boesch, 2003; Hobaiter and Byrne, 2012; Cäsar et al., 2013). Thus, research bears out that, while monkeys, apes, and humans share the cognitive abilities that are required to learn and use signs, only humans develop vast systems of symbols, and they do so primarily in a vocal medium. One implication of these results is that, even though cognitive abilities can be essential prerequisites in acquiring and manipulating symbols, some other capacity is needed to account for the emergence of these types of signs in vocal communication.

## The Role of the Medium in the Rise of Symbolic Communication

### The Case Against Modality-Independent Accounts: “Sign Languages” and Storylines of the Gestural Origin of Language

Compared to cognitive skills, the capacity to articulate vocal patterns stands as an obvious human-specific trait. Yet in the literature on language origin, many researchers are guided by the belief that language emerged in a gestural medium. Several findings have motivated this view, which implies a conceptualization of “language” as an amodal function. One

pivotal finding relates to the lack of voluntary control of vocalization in nonhuman primates.

In particular, studies by Jürgens (1976, 1992) showed that the brains of monkeys and apes lack monosynaptic fibers linking the motor cortex to laryngeal-muscle motoneurons in the nucleus ambiguus. Such findings concur with the poor control over reactive vocalizations in these species (Jürgens, 2002, pp. 242–245; Simonyan and Horwitz, 2011; Simonyan, 2014). The nervous systems of monkeys and apes do, however, present direct monosynaptic projections to motoneurons associated with the control of finger muscles, and to jaw- and lip-muscle motoneurons in the trigeminal and facial nuclei (Jürgens, 2002; Sasaki et al., 2004; Lemon, 2008). Compared to humans, though, there are fewer direct connections to tongue-muscle motoneurons in the hypoglossal nucleus (Kuypers, 1958; Jürgens, 2002), which accords with the paucity of oral segmentations or syllable-like patterns in the calls of non-human primates (Lieberman, 1968, 2006a; Lieberman et al., 1992). Taken together, these observations may have led many to believe that, since apes and monkeys produce vocal signals as inflexible reactions, symbolic language instead evolved from controllable hand gestures (e.g., Hewes, 1996; Corballis, 2002, 2009, 2017; Tomasello and Zuberbühler, 2002; Arbib, 2005, 2012, 2015; Gentilucci and Corballis, 2006; Pollick and de Waal, 2007; Arbib et al., 2008). This held belief, however, conflicts with the general observation that gestural signs of non-human primates do not function *symbolically* (and are mostly iconic or indexical). Thus, theories of the gestural origin of language are not supported by observations of extant species. Instead, the theories refer to *indirect* evidence seen to support storylines which essentially suggest that the last common ancestor (LCA) of the *homo* and *pan* (chimpanzee and bonobo) genera had, at some point, developed *symbolic gestures* from which spoken language evolved. This view, popularized by authors like Corballis and Arbib (as outlined below), has been criticized on fundamental grounds.

One objection bears on the theoretical significance given to “sign languages.” Proponents of gestural theories frequently refer to sign languages as illustrating the possibility of a gestural stage in language evolution (e.g., Corballis, 2002; Arbib, 2012, chapter 6). As such, this interpretation adheres to a view of language as deriving from an amodal faculty (e.g., Chomsky, 1965, 2000, 2006; and also Pinker, 1994; Hauser et al., 2002; Hauser, 2016). For example, gestural signs are seen to support the idea that “the language faculty is not tied to specific sensory modalities” (Chomsky, 2000, p. 121), and that “discoveries about sign languages [...] provide substantial evidence that externalization is modality independent” (Chomsky, 2006, p. 22). However, such claims repeatedly disregard the fact that there is no known case where a community of normal hearers develops a sign language as a primary means of communication (Emmorey, 2005). Said differently, gestural signs as a primary system of communication generally appear where people share a pathology affecting the hearing modality – which hardly supports the notion of an amodal language capacity. On the contrary, it suggests that, given the normal human ability to control *both* vocalization and hand gestures, symbolic communication links to a vocal-auditory

modality with visual gestures having an accessory imagistic role (McNeill, 2005). It follows that an account of the rise of spoken language requires an explanation of how and why symbolic signs link to the vocal medium. But assuming the gestural origin of language leads instead to posit an evolutionary shift from gestural to vocal signs, which presents a conundrum for evolutionists.

To explain this briefly, one can refer to the theories of Corballis (1992, 2003, 2009, 2010, 2017; Gentilucci and Corballis, 2006), and Arbib (2005, 2011, 2012, 2015, 2017; Arbib et al., 2008). A critical claim of these proposals is that a left-lateralized control of hand gestures in area F5 of the monkey cortex evolved into a left-sided dominance for language, which the authors locate in Broca's area. Corballis and Arbib also refer to research showing that mirror neurons in F5 discharge when a monkey observes hand motions in others (see, e.g., Rizzolatti and Arbib, 1998; Ferrari et al., 2005). Both see in these responses a mechanism of action understanding, and conjecture that mirror neurons played a role in the shift from a gestural to a vocal-modality of communication. On how this shift occurred, it is speculated that, when the LCA descended from trees and developed bipedalism, there was a freeing of the hands allowing the development of expressive manual signs. At first, this led to putative iconic pantomimes that became conventionalized and symbolic (at a "protosign" stage), before "protospeech" developed. However, it is difficult to find in this narrative a working mechanism that converts gestures to vocal signs. For example, Arbib recently explained that pantomimes created an "open-ended semantics" which "...provides the adaptive pressure for increased control over the vocal apparatus" (Arbib, 2015, pp. 612–613; see also 2017, p. 144). By this account, the semantics of protosigns "...establishes the machinery that can then begin to develop protospeech, perhaps initially through the utility of creating novel sounds to match degrees of freedom of manual gestures (rising pitch could represent an upward movement of the hand)" (2015, p. 613). Thus, the core explanation in this view is that *semantics* drove the evolution of vocalization, and the pairing of physiological parameters of hand control with those of articulatory, laryngeal, and respiratory systems of (proto-) speech. In this account, the example of iconic signs (of pitch increase and hand rising) hardly helps to understand how vast systems of *symbolic* signs emerged. Critics have also questioned whether any realistic model can be devised to "translate" hand motions into sequences of vocalized sounds (Hewes, 1996; MacNeilage, 2008/2010, pp. 287–288).

In weighing the above scenario, one should note that it rests on the claim that a left-sided control of hand motions in the monkey cortex overlaps mirror neurons and Broca's area in humans. It has been reported, however, that activity in mirror neurons during the perception and production of hand motions is not left-lateralized (Aziz-Zadeh et al., 2006). More generally, one might question the *a priori* validity of theories where semantics is seen to drive the evolution of mechanisms of vocalization. Such views are not limited to gestural accounts. They extend to a variety of theories that focus on cognitive skills while overlooking modality-specific constraints and how they shape signals and signs. In this orientation, it is as if symbol systems generate from some amodal mental function.

For example, Deacon (1997, 2012) suggests an account whereby signs created by apes (and children) first appear to have an indexical function. Then, when the signs are logically combined, they become more symbolic (Deacon, 1997, chapter 5). But again, apes in natural environments do not develop symbolic communication despite a capacity to combine given signs (as per the experiments of Savage-Rumbaugh et al., among others), so the question remains: how is it that symbolic signs emerge for humans and why is this linked primarily to the vocal medium? On these questions, Deacon (1997) basically offers a circular explanation: "...language must be viewed as its own prime mover. It is the author of a co-evolved complex of adaptations arrayed around a single core semiotic innovation. . ." (p. 44).

Aside from core semantics or semiotic functions, some theories also submit that the evolution of symbolic communication was driven by socio-cognitive functions. For example, the theory of "interactional instinct" suggests that, in language acquisition as in language evolution, children signal their intention to do something, and "The intent becomes a symbol that the child expresses in an emotionally based interaction. . ." (Lee and Schumann, 2005, p. 7; Lee et al., 2009). In related proposals, constellations of mental skills including "shared intentionality," "perspective-taking," "comprehension" (etc.), along with "thought processes" (Ackermann et al., 2014), and "purpose" (Deacon, 2011) are evoked as driving factors (for an overview of these types of factors, see the proposal of Pleyer, 2017, and Pleyer and Winters, 2015). These accounts collectively imply what some have called "mentalist teleological" principles (Allen, 1996/2009), which do not accord with accepted features of evolution. In particular, one feature holds that evolution reflects biological change in relation to physical aspects of the environment (see Christiansen and Chater, 2008). Accepting this, it is difficult to fathom how the evolution of anatomical structures of vocal communication would be driven by "semantics," "thoughts," "intentions" (etc.), without some basis in the sensory effects of physical signals and physiological processes of sign production. Indeed, if one *defines* symbols as entailing associations between concepts and signal elements, then the notion that symbolic language arose from amodal mental factors presents a contradiction in terms in that, in the absence of a modality of communication, there are no signals to which concepts can associate (but cf. the pronouncement that language has little to do with communication: Chomsky, 2002, pp. 76–77; 2012, p. 11).

## Modality-Dependent Accounts of the Rise of Symbolic Language

### *Mimesis and procedural learning*

Contrary to the above viewpoint, several authors have submitted that language first evolved in the vocal modality and that this medium imposes particular constraints on the formation of signs (e.g., Lieberman, 1984, 2000, 2003, 2007, 2016; MacNeilage, 1998, 2008/2010; Studdert-Kennedy, 1998; MacNeilage and Davis, 2000, 2015). This basically refutes claims that sign languages and spoken languages have similar structure reflecting an amodal mental capacity (e.g., Bellugi et al., 1989). Thus,

MacNeilage (2008/2010, chapter 13) explained that vocal and gestural modalities shape signs quite differently: for instance, gestural signs are holistic and can involve simultaneous hand and body motions whereas, in spoken language, sounds are strictly sequential, and are constrained by articulatory and respiratory-phonatory systems. It is useful to note that such a viewpoint is echoed in recent studies of emerging sign language that explicitly focus on the link between signs and constraints on gestures. Most revealing is the work of Sandler et al. on *Al-Sayyid Bedouin Sign Language* (ABSL), which led to expressed reservations both on gestural accounts of language origin, and linguistic methods which have dominated research on sign languages.

Traditionally, the study of sign languages has relied on linguistic analyses and assumptions which focus on abstract phonological features, “word” categories and their syntax, but which entirely neglect modality-specific differences between gesture and sound production (MacNeilage, 2008/2010). Sandler et al. noted several difficulties in attempting to pigeon-hole gestures in terms of assumed linguistic features and units, and instead examined gestures as such (Sandler, 2012, 2013, 2017; Sandler et al., 2014). Their study of ABSL revealed a correspondence between developing motor aspects of gestures and the complexity of expressed concepts, which question the doctrine that spoken and gestural systems can derive from a common modality-independent function. As Sandler (2013, pp. 194–195) remarked, in commenting Arbib’s gesture theory of language origin –

...a different motor system controls language in each modality [gestural and vocal], and the relation between that system and the grammar is different as well. Considering the fundamental differences in motor systems, I am mindful of the reasoning of experts in the relation between motor control and cognition (Donald, 1991; MacNeilage, 2008/2010) who insist on the importance of the evolution of the supporting motor system in the evolution of language to the extent that ‘mental representation cannot be fully understood without consideration of activities available to the body for building such representations... [including the] dynamic characteristics of the production mechanism’ (Davis et al., 2002).

The findings also implied a reevaluation of linguistic assumptions on sign languages (Sandler, 2012, pp. 35–36):

...many believe that sign languages rapidly develop into a system that is very similar to that of spoken languages. Research charting the development of Nicaraguan Sign Language (e.g., Senghas, 1995; Kegl et al., 1999), which arose in a school beginning in the late 1970s, has convinced some linguists, such as Steven Pinker, that the language was “created in one leap when the younger children were exposed to the pidgin signing of the older children...” (Pinker, 1994: 36). This is not surprising if, as Chomsky believes, “...language evolved, and is designed, primarily as an instrument of thought, with externalization a secondary process.” (Chomsky, 2006: 22). [...] Our work on ABSL, of which the present study is a part, suggests that those of us who contributed to this general picture may have overstated our case.

In sum, motor aspects of gesture and sound production shape symbol systems differently and storylines suggesting that vocal

symbols developed from gestural signs bear intractable problems of gesture-to-sound conversion. But while such problems suggest a necessary link between symbolic language and vocal processes, not all modality-specific accounts deal with the rise of symbolic signs.

As an example, Lieberman’s proposal focuses on how spoken language, viewed as a combinatorial system, links to the evolution of speech processes in conjunction with the basal ganglia and its function as a “sequencing engine” (e.g., Lieberman, 2003, 2006b, 2016). However, this proposal does not specifically address the issue of how vocal symbols emerged. MacNeilage, on the other hand, views symbols as the most fundamental factor of language evolution (MacNeilage, 2008/2010, p. 137). In his “frame/content” theory, the vocal modality provided a prototypical frame, the syllable, which is seen to originate in cyclical motions such as mastication (MacNeilage, 1998, but see Boucher, 2007 who shows that cyclicity in speech articulation has little to do with mastication or chewing motions). MacNeilage argues that, for arbitrary symbols to emerge, “...hominids needed to evolve a capacity to invent expressive conventions” (p. 99), though this did not arise from a “higher-order word-making capacity” (as in Hauser et al., 2002). Instead, the capacity to form “words” rests on mimesis and procedural learning, which refines actions and memory of actions. Donald (1997) described these processes as a capacity to “...Rehearse the action, observe its consequences, remember these, then alter the form of the original act, varying one or more parameters, dictated by the memory of the consequences of the previous action, or by an idealized image of the outcome” (p. 142). MacNeilage, quoting Donald, submits that while human infants manifest procedural learning at the babbling stage, great apes do not, so that “It would be no exaggeration to say that this capacity is uniquely human, and forms the background for the whole of human culture including language” (Donald, p. 142). Yet studies of tool manufacture in chimpanzees challenge this claim.

For instance, in a recent longitudinal study, Vale et al. (2016, 2017) observed that chimpanzees are not only successful in creating tools to retrieve rewards. They can also retain a procedure of tool manufacture for years and transfer this knowledge to new tasks, indicating an ability to acquire skills. Again, procedural learning, as other cognitive abilities, is required in symbol learning. However, it is not a sufficient factor in accounting for the human-specific development of symbolic language in a vocal medium. Nonetheless, the above proposals share the view that symbolic communication does not emerge from mental functions alone but essentially links to a capacity to produce patterns like syllables and babble.

#### **“Sound symbolism”: Questions of efficiency and ease of learning of iconic signs**

Another theory which also rests on mimesis suggests that symbolic language may originate from iconic vocal signs (or sound symbols) on the assumption that “iconicity seems easier” than making arbitrary associations (Kita, 1997; Monaghan et al., 2012; Thompson et al., 2012; Imai and Kita, 2014; Sereno, 2014; Perlman et al., 2015; Massaro and Perlman, 2017). The idea that symbols arose from a mimicking of objects and events partly

draws from experiments by Sapir (1929) and Köhler (1947) on “sound-shape” pairings, where listeners judge perceived consonants and vowels as relating, for instance, to “angular” and “rounded” forms, or Ohala’s (1994) “frequency code” where pitch is related to features like “size” and “brilliance” (e.g., Maurer et al., 2006; the frequency code also extends across species that use vocal pitch in signaling aggressive/passive intentions; Morton, 1994). The central assumption is that such iconic cues are vital to the ontogenesis and phylogenesis of language because they inherently facilitate the sound-meaning mappings and the displacement of signs (e.g., Monaghan et al., 2014). Experiments showing sound-form associations in infants and adult learners are often cited as supporting this view. For instance, a study by Walker et al. (2010) shows that pre-babbling infants of four months are able to associate pitch to features such as height and brightness (see also Monaghan et al., 2012). However, the experiments do not serve to demonstrate the facilitating effect of iconic sounds on language development, or the necessity of an iconic stage in developing arbitrary signs of oral language.

On these issues, attempts to relate iconic signs to spoken language face a logical problem. Iconic gestures or sounds offer highly restricted sets of signs compared to arbitrary symbols. Any restriction on the number of signs will inherently limit the diversity of form-referent associations and thus the *efficiency* of signs in communicating fine distinctions in meaning, leading some to see that “While sound symbolism might be useful, it could impede word learning. . .” (Imai and Kita, 2014, p. 3; also Monaghan and Christiansen, 2006; Monaghan et al., 2012). This is evident in signs serving to *name* referents where sound symbols are quite limited (e.g., it is difficult to conceive how one could mimic open sets of lexemes such as “proper nouns”). On the idea that vocal iconic signs may nonetheless be easier to learn than arbitrary signs, computational models and experiments involving adult learners lead to the opposite conclusion (Gasser, 2004; Monaghan et al., 2011, 2012, 2014). For instance, in a series of experiments, Monaghan et al. (2012, 2014) compared the learning of arbitrary and iconic (systematic) sign-meaning pairs taking into account given co-occurring “contextual” elements. In both neural network simulations and behavioral experiments, arbitrary form-meaning pairs were learned with fewer errors and more rapidly than iconic pairs. Still, such tests do not address the issue of how *arbitrary* signs arise in the vocal medium, and why this is specific to humans.

### ***Segmented vocalization and the rise of symbolic signs: A demonstration***

The aforementioned simulations and tests focus on the learning of sound-shape associations by infants and adults with respect to the perception of *given* signs provided by an experimenter. Such protocols do not address the issue of how signs emerge, and infants in the first months of life may not produce symbolic signs. However, it is well-established that maturational changes in the vocal apparatus coincide with the rise of such signs in child speech. On this development, comparisons of the vocal processes of human infants and apes are revealing of the mechanisms underlying the rise of vocal symbols. In particular, pre-babbling infants, like nonhuman primates, are obligate nasal breathers and

produce sounds with nasal resonances (Negus, 1949; Crelin, 1987; Thom et al., 2006). Acoustically, nasalization dampens upper harmonics and formants, reducing the distinctiveness of sounds. Moreover, continuous nasal air-flow during vocalization implies that articulatory motions may not create salient features of oral segmentation (Lieberman, 1968, 1984; Lieberman et al., 1972; Thom et al., 2006). For instance, producing multiple elements like stops [p,t,k,b,d,g], fricatives [f,s,ʃ,v,z,ʒ], or any articulatory pattern of segmentation requires modulations of oral pressure which are difficult to achieve in a system where air flows through the nose. For this reason, early productions of infants largely appear as continuous nasalized cries and vocalizations which, as in the vocal sounds of non-human primates, are divided by breath interruptions and glottal closures (Crelin, 1987; Lynch et al., 1995). At three months, though, humans manifest a control of pitch contours, which can reflect a refinement of monosynaptic connections between the motor-cortex and motoneurons of laryngeal muscles (Ackermann et al., 2014). Some suggest that, at this stage, vocalizations become less reactive and iconic, and can involve a symbolic coding of pitch (Oller, 2000; Oller and Griebel, 2014). Subsequent supraglottal changes occur which are also human specific.

Of interest is the progressive decoupling of the nasopharynx indexed by the distancing of the epiglottis from the velum. In human newborns, as in many mammals, the epiglottis habitually overlaps the velum creating a sealed passage that allows nasal breathing while ingesting food and extracting breast milk (Negus, 1949; Kent, 1984; Crelin, 1987). It should be noted that the decoupling of the nasopharynx involves *soft* tissues – hence the difficulty of dating this change from fossil records. Many works of comparative anatomy involving CT and MRI scans discuss this decoupling by reference to a “descent of the larynx” which is indexed by a lowering of the hyoid bone attached to the larynx, and related measures of vocal-tract length. Because this lowering is observed across species, some have concluded that laryngeal descent has little to do the rise of spoken language in humans (e.g., Fitch and Reby, 2001; Fitch, 2010, chapter 8). However, laryngeal descent is only accompanied by a *permanent* decoupling of the epiglottis and the velum in humans, beginning at about 6–8 months (Kent, 1984; Nishimura et al., 2006; Lieberman, 2007). Following this decoupling, articulatory motions can create a variety of salient oral features in vocal signals, which constitutes a pivotal human-specific development. Though some contend that other species can produce a range of vowel-like resonances (Fitch, 2000, 2005; Fitch and Reby, 2001), only humans segment these resonances using varying articulatory motions. This important change is accompanied by a general increase in rhythmic behavior (Iverson et al., 2007) leading to *babble* (Locke, 1995; Oller, 2000). Combining this morphological development with a capacity to modulate pitch contours confers the unique ability to manipulate both tonal and articulatory patterns. As noted, compared to other primates, humans develop direct monosynaptic projections to motoneurons of laryngeal muscles and also present a greater ratio of direct connections to tongue-muscle motoneurons which together support a control of orally modulated patterns of vocalization.

The symbolic potential of these patterns can be fathomed by considering that, in canonical babble, reduplicated syllables containing articulatory features such as [dada], [mamama], [nana] (etc.) become rapidly associated with caregivers, food, and other contextual referents. These early signs show that the *arbitrariness* underlying symbolic language can be *inherent* to the types of sounds that arise with orally segmented speech (a point also noted by Corballis, 2002), and that contextual information suffices to establish functional sound-meaning associations. In fact, the developmental literature does not suggest that iconic sounds or gestural mimics precede the rise of symbolic signs in canonical and variegated babble (Locke, 1995; Oller, 2000). On the other hand, it is generally acknowledged that the rise of vocal symbols accompanies a shift from pitch-varying to orally segmented vocalizations, though some symbol coding of pitch can precede this shift (Oller and Lynch, 1992; Oller, 2000).

Overall, the above developments suggest that the human capacity to produce orally segmented vocalizations presents a *necessary* – though not sufficient – factor in the rise of symbolic signs. In other words, cognitive and perceptual abilities are certainly required in forming sound-meaning associations, but these abilities are present at a basic level in non-human primates who do not develop symbolic signs like those of spoken language. Hence, the particular shift from pitch-varying to articulated vocalizations appears essential to the emergence of efficient vocal symbols in that, compared to pitch patterns, oral modulations offer a more diverse set of salient articulatory features by which to create fine distinctions in meaning.

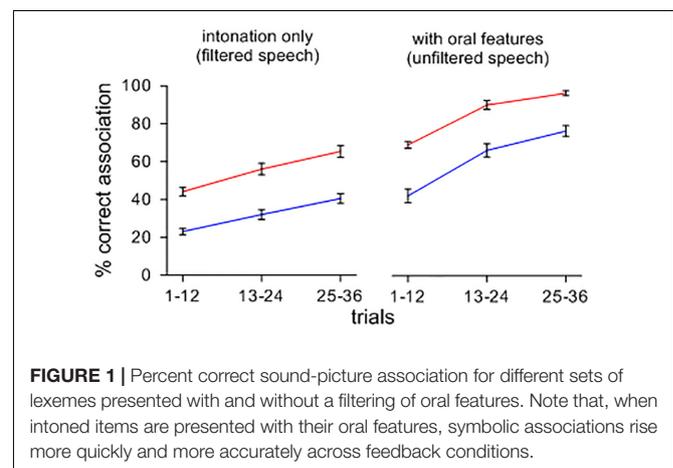
Of course, a direct demonstration of this shift on the development of vocal symbols is not possible (i.e., one may not experimentally manipulate the human ability to articulate sounds). But one can artificially reduce acoustic features associated with articulatory motions by filtering signals so as to observe effects on the formation of symbolic associations. This guided the design of a straightforward demonstration using a task where listeners learn to associate pictures to unfamiliar speech sounds (in this case, two-syllable “words” spoken in Mandarin). In the task, sounds are presented with sets of possible referents (pictures of familiar objects) allowing the rise of associations through repeated trials. To evaluate the effect of a shift from intoned to orally articulated patterns on symbolic associations, filtered stimuli are presented bearing only pitch-varying intonations, then unfiltered stimuli with oral features are presented (for an example of the stimuli, see “Materials and methods” **Figure 3**). To support word-picture associations, two types of feedback are used representing two basic responses that can be obtained in a context of verbal learning. In the first type (henceforth *feedback A*), learners make an association and obtain an indication on whether or not the association is correct (“yes/no”). In the second type (*feedback B*), learners additionally receive information on what the sound designates (the correct picture is displayed). The first feedback condition emphasizes a process of inference, while the second favors rote learning. In both conditions, the prediction was that symbolic sound-meaning associations would form more rapidly for orally segmented sounds than for intoned patterns principally because of the

greater efficiency of oral features in making distinctions in meaning.

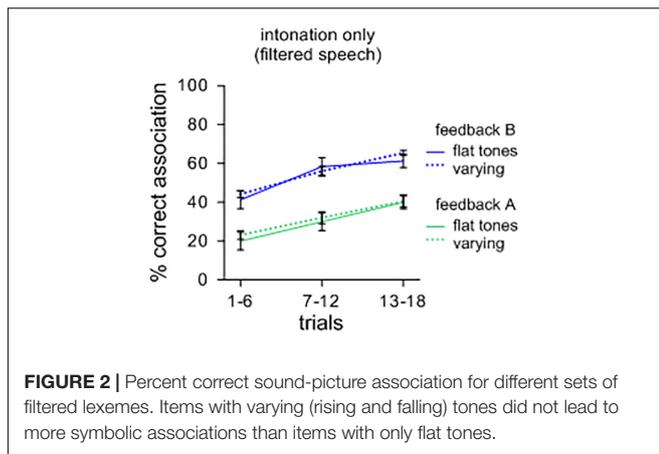
## RESULTS

**Figure 1** shows the effect of a change from intoned to orally modulated patterns on the rise of symbolic associations. Upon hearing sets of items containing features of oral articulation, arbitrary sound-meaning associations shift and form more rapidly and accurately than when the items are heard with only their tonal attributes, and this effect appears across feedback conditions. A repeated-measures ANOVA confirmed main effects of stimuli type [ $F(1,39) = 2.202$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.822$ ;  $MSE = 0.012$ ], and feedback ( $[F(1,39) = 3.581$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.896$ ;  $MSE = 0.011$ ]). Orally segmented stimuli yielded more correct associations than filtered stimuli, and *feedback B* led to more correct associations than *feedback A*. There was no significant interaction between the type of contexts and feedback conditions [ $F(1,39) = 8.740E-5$ ; n.s.;  $\eta_p^2 = 0.000$ ;  $MSE = 0.012$ ].

As a further illustration, a separate comparison of sound-meaning associations was performed for sets of filtered items with different tones of Mandarin. Half of the contexts had initial flat tones while the other half contained variable rising and falling tones. One could surmise that symbolic associations could be facilitated for sets of items with varying tones compared to items that have only flat tones. However, as **Figure 2** illustrates, this effect did not occur. A repeated measures ANOVA showed no main effects of tonal patterns [ $F(1,39) = 0.049$ ; n.s.;  $\eta_p^2 = 0.069$ ;  $MSE = 0.017$ ], but a main effect of feedback [ $F(1,39) = 2.175$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.757$ ;  $MSE = 0.018$ ] with *feedback B* leading to more correct associations than *feedback A*. Again, there was no significant interaction between the types of contexts and feedback [ $F(1,39) = 0.000$ ; n.s.;  $\eta_p^2 = 0.001$ ;  $MSE = 0.013$ ]. A visual comparison with **Figure 2** suggests that, overall, intonational features are less efficient in supporting the formation of symbolic associations than oral features. It should also be weighed that many languages unlike Mandarin do not code tone at all in distinguishing lexical items, and thus, from a general perspective,



**FIGURE 1 |** Percent correct sound-picture association for different sets of lexemes presented with and without a filtering of oral features. Note that, when intoned items are presented with their oral features, symbolic associations rise more quickly and more accurately across feedback conditions.



intonation as such may not suffice in distinguishing large repertoires of lexemes.

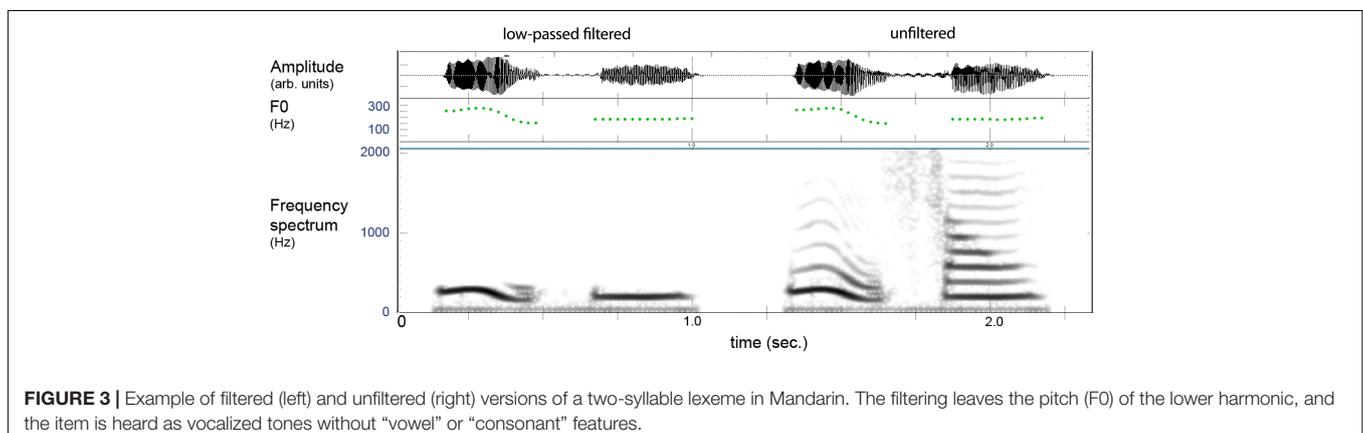
## DISCUSSION

The preceding results demonstrate a seemingly self-evident effect: beyond the competence of the above learners and their experience with language, they were more prone to form arbitrary sound-meaning associations with oral patterns of speech than when speech contained only intonational patterns. Specifically, when shifting from items with tonal features to items that include features of articulation, symbolic associations were formed more rapidly and accurately. Of course, these observations and the above test are not meant to reflect evolution. Their purpose is to demonstrate the basic point that one may not account for the rise of symbolic signs in language without some reference to the types of signals that a modality of expression affords. Yet this is not a generally accepted principle.

As seen in the above review, the idea that symbolic language evolved from modality-independent cognitive abilities is widespread. One should also remark that the definitions of symbols that guide much of the work on language evolution refer to 19th century writers like Peirce and Saussure who did not consider how processes of modality can shape signals and

signs. Following this tradition, storylines of language origin are largely oriented by the belief, popularized in linguistic theory, that symbolic language derives from a mental competence or faculty which has little to do with processes of expression like gestures and speech (Saussure, 1916/1966; Chomsky, 1965). From this standpoint, the rise of symbolic language would seem “mysterious” (Hauser et al., 2014), “puzzling” (Aitchison, 2000; Bouchard, 2015), and can lead to ask “why only us” (Berwick and Chomsky, 2015) or to speculations of evolutionary saltations (e.g., Hauser et al., 2002). However, linguistic analyses that focus on abstract categories and units overlook the structural effects of varying modalities of expression. On these effects, we noted that work on emerging sign and spoken language bears out that motor aspects of gestures and vocalization do shape sign systems very differently. As for the idea that symbols arise from a mental capacity, decades of research has shown that monkeys and apes have a basic competence to learn, process, and combine symbolic signs. Yet they do not develop the types of productive vocal symbols found in spoken language. One implication is that the human specificity of symbolic language may not be explained in terms of cognitive capacities alone. Other factors relating to the human ability to control vocal signals are needed to account for symbolic language and why it arises primarily in oral medium.

On these factors, several works of comparative physiology have identified human-specific changes that can underlie the rise of symbolic signs, and these reflect in ontogenesis in terms of a shift from pitch-varying patterns to orally articulated babble. As Corballis (2002), pp. 187–188 indicated, articulatory modulations of sounds generate numerous features in signals creating inherently arbitrary signs. These signs can be rapidly associated with co-occurring referents in the course of developing language. No stage of iconic gestures or “sound-symbolism” appears to precede this development (Locke, 1995; Oller, 2000). But nor does such a stage seem necessary. In fact, several studies have shown that infants, even pre-babblers, readily associate heard “words” with referents, and use these symbols to categorize objects or parts of objects (see Fulkerson and Waxman, 2007; Ferry et al., 2010, and for children of 12 months, MacKenzie et al., 2011). Interestingly, the reports also show that infants are less successful in forming symbolic associations when presented with sounds like intoned [mmm] or sounds that cannot be articulated,



and instead attend to familiar speech (MacKenzie et al., 2011; Marno et al., 2015, 2016; and for evidence that activity in language areas of the brain are organized in terms of speech sounds, see Magrassi et al., 2015). This ability to acquire symbols is not distinctly human, as we noted, but communication by way of orally articulated signs is. One can only speculate on what might have evolved had humans been limited to pitch-varying calls. From the above results, it can be surmised that pitch-controlled signals without features of oral articulation would restrict the rapid formation of efficient (accurate) symbols. On the phylogenesis of this capacity, certainly a pivotal factor is the decoupling of the nasopharynx which contributed to free the oral tract allowing an articulatory modulation of sounds (e.g., Negus, 1949; Lieberman, 1984, 2006b; Crelin, 1987). Some see this as a consequence of bipedalism (originating 5–7 million years ago) and that bipedalism further led to a decoupling of respiration and locomotion that supported an independent control of phonation (Provine, 2005; see also Maclarnon and Hewitt, 1999, 2004). However, as we mentioned, bone markers in fossil records may not serve to index the separation of the nasopharynx, so dating this change as co-evolving with bipedalism appears problematic.

## MATERIALS AND METHODS

### Participants

The participants were 40 speakers aged 20 to 41 years (mean of 25 years; 19 females) with no history of hearing problems or knowledge of any tone language. Prior to the task, all were screened for normal memory performance using a standard digit-span test (Wechsler, 1997).

### Speech and Picture Stimuli

The original speech stimuli consisted of 24 two-syllable Mandarin lexical items (“words”) that matched 24 black-and-white pictures of objects (Snodgrass and Vanderwart, 1980). Half of these contexts had an initial flat tone and half a varying rising or falling tone (used to compare effects of coded pitch). The items were produced in isolation by a native female speaker of Mandarin and recorded using a headset microphone (AKG, model C-477-WR) and an external sound card (Shure, model X2u) set to a 16-bit resolution and 44.1 kHz sampling rate. The recorded items were stored as separate .wav files, and copies of these files were low-pass filtered at 350 Hz using an IIR function (*Goldwave*, v. 6.19) giving 24 filtered and 24 unfiltered versions of each of the lexemes. Finally, all of the stimuli were amplitude normalized to obtain similar dB intensities.

**Figure 3** provides an example of the filtered and unfiltered version of a spoken lexeme where one can see that only the lower harmonic is present in the filtered version, which gave natural-sounding pitch patterns with a vocalic quality. The distinctiveness of these contexts was evaluated by an external judge (not the authors) using a discrimination test. This was done by creating duplicates (*AA*) of the filtered items and placing another filtered test item (*X*) with similar tones before and following the pair, giving series of *AAX* and *XAA*. All tokens *X*

were correctly discriminated indicating that the filtered stimuli retained distinguishable elements.

### Test Design and Procedure

A repeated measures design was used with a counterbalancing of blocks of spoken lexemes on two type conditions (filtered vs. unfiltered), and two feedback conditions (*A* vs. *B*, described earlier), while taking into account a specific order of presentation (filtered items with *feedback A* were presented before filtered items with *feedback B*; then, unfiltered items with *feedback A* were presented before unfiltered items with *feedback B*). The latter presentation sequence basically aimed to capture effects of a shift from tonal to orally segmented patterns. The 24 lexemes were arranged into four lists of trials, each containing six repetitions of six different lexemes, giving 36 test trials per list. Each trial lexeme within the lists was matched to an array of four pictures with one correct lexeme-picture match and three random fillers. The counterbalance design implied that, across participants, all lists were presented an equal number of times in each condition.

In the experiment, a test trial began with a three-second display of four pictures, followed by a heard speech stimulus. The participant then had to select a likely speech-picture match by pressing a number on a keyboard (a forced-choice task). After the participant’s key-press response, a feedback slide was displayed for two seconds. During the test, participants sat in a sound-treated room in front of a monitor and listened with headphones (Beyerdynamic, model DT 250) to the speech stimuli which were delivered via software (*E-Prime 2*). The sound stimuli had peak values of 71 dBA at the ears, as measured with a sound-level meter and headphone adapter (Digital Recordings, model DR-1). Practice runs were provided prior to the test, which proceeded without interruption. Statistical analyses of the responses used the procedures of *SPSS* (v. 17.0).

### ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ‘Declaration of Helsinki’ with written informed consent from all subjects. The protocol was approved by the Comité d’Éthique de la Recherche de la Faculté d’Arts et Sciences de l’Université de Montréal.

### AUTHOR CONTRIBUTIONS

VB conceived the theoretical arguments and, with AG, designed the test. AR-B contributed to the data acquisition and procedure. All three authors contributed to the data analysis, interpretation, critical revision, and final approval of the paper.

### FUNDING

The research was supported in part by a grant of the *Fonds Québécois de la Recherche sur la Nature et la Technologie* (FQRNT No. 175811).

## REFERENCES

- Ackermann, H., Hage, S. R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci.* 37, 529–546. doi: 10.1017/S0140525X13003099
- Aitchison, J. (2000). *The Seeds of Speech: Language Origin and Evolution*. Cambridge: Cambridge University Press.
- Allen, C. (1996/2009). “Teleological notions in biology,” in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Stanford CA: Stanford University).
- Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–124. doi: 10.1017/S0140525X05000038
- Arbib, M. A. (2011). From mirror neurons to complex imitation in the evolution of language and tool use. *Annu. Rev. Anthropol.* 40, 257–273. doi: 10.1146/annurev-anthro-081309-145722
- Arbib, M. A. (2012). *How the Brain got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199896684.001.0001
- Arbib, M. A. (2015). “Language evolution,” in *The Handbook of Language Emergence*, eds B. MacWhinney and W. O’Grady (Hoboken, NJ: John Wiley & Sons), 600–623.
- Arbib, M. A. (2017). Toward the language-ready brain: biological evolution and primate comparisons. *Psychon. Bull. Rev.* 24, 142–150. doi: 10.3758/s13423-016-1098-2
- Arbib, M. A., Liebal, K., and Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* 49, 1053–1076. doi: 10.1086/593015
- Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., and Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *J. Neurosci.* 26, 2964–2970. doi: 10.1523/jneurosci.2921-05.2006
- Belin, P. (2006). Voice processing in human and non-human primates. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2091–2107. doi: 10.1098/rstb.2006.1933
- Bellugi, U., Poizner, H., and Klima, E. S. (1989). Language, modality and the brain. *Trends Neurosci.* 12, 380–388. doi: 10.1016/0166-2236(89)90076-3
- Berwick, R. C., and Chomsky, N. (2015). *Why Only us: Language and Evolution*. Cambridge MA: MIT Press.
- Bolhuis, J. J., Tattersal, I., Chomsky, N., and Berwick, R. C. (2012). How could language have evolved? *PLoS Biol.* 12:e1001934. doi: 10.1371/journal.pbio.1001934
- Bouchard, D. (2015). Brain readiness and the nature of language. *Front. Psychol.* 6:1376. doi: 10.3389/fpsyg.2015.01376
- Boucher, V. J. (2007). “Serial-order control and grouping in speech: findings for a frame/content theory,” in *Proceedings of the XVth International Congress of Phonetic Sciences*, eds J. Trouvain and W. J. Barry (Saarbrücken: ICPhS), 413–416.
- Bullock, M. J., Boysen, S. T., and Furlong, E. E. (2008). Visual attention and its relation to knowledge states in chimpanzees, *Pan troglodytes*. *Anim. Behav.* 76, 1147–1155. doi: 10.1016/j.anbehav.2008.01.033
- Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192. doi: 10.1016/j.tics.2008.02.010
- Cäsar, C., Zuberbühler, K., Young, R. J., and Byrne, R. W. (2013). Titi monkey call sequences vary with predator location and type. *Biol. Lett.* 9, 1–5. doi: 10.1098/rsbl.2013.0535
- Cheney, D. L., and Seyfarth, R. M. (1990). *How Monkeys see the World*. Chicago, IL: University of Chicago Press.
- Chomsky, N. (1965). *Aspects of the theory of Syntax*. Cambridge MA: MIT Press.
- Chomsky, N. (1972). *Language and Mind*. New York, NY: Harcourt Brace Jovanovich.
- Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511811937
- Chomsky, N. (2002). *On Nature and Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511613876
- Chomsky, N. (2006). *Language and mind*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791222
- Chomsky, N. (2012). *The Science of Language. Interviews with James McGilvray*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139061018
- Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–558. doi: 10.1017/S0140525X08004998
- Christiansen, M. H., and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 300–307. doi: 10.1016/s1364-6613(03)00136-0
- Clark, B. (2012). Syntactic theory and the evolution of syntax. *Biolinguistics* 7, 169–197.
- Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition* 44, 197–226. doi: 10.1016/0010-0277(92)90001-X
- Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton NJ: Princeton University Press.
- Corballis, M. C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav. Brain Sci.* 26, 199–208; discussion 208–160. doi: 10.1017/S0140525X03000062
- Corballis, M. C. (2009). The evolution of language. *Ann. N. Y. Acad. Sci.* 1156, 19–43. doi: 10.1111/j.1749-6632.2009.04423.x
- Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain Lang.* 112, 25–35. doi: 10.1016/j.bandl.2009.02.002
- Corballis, M. C. (2017). *The Truth about Language: What it is and Where it Came from*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226287225.001.0001
- Crelin, E. S. (1987). *The Human Vocal Tract: Anatomy, Function, Development and Evolution*. New York, NY: Vantage Press.
- Crockford, C., and Boesch, C. (2003). Context-specific calls in wild chimpanzees, *Pan troglodytes* verus: analysis of barks. *Anim. Behav.* 66, 115–125. doi: 10.1006/anbe.2003.2166
- Davis, B. L., MacNeilage, P. F., and Matyear, C. L. (2002). Acquisition of serial complexity in speech production: a comparison of phonetic and phonological approaches to first word production. *Phonetica* 59, 75–107. doi: 10.1159/000066065
- Deacon, T. W. (1997). *The Symbolic Species*. New York, NY: W.W. Norton.
- Deacon, T. W. (2011). *Incomplete Nature: How mind Emerged from Matter*. New York, NY: W.W. Norton.
- Deacon, T. W. (2012). “Beyond the symbolic species,” in *The Symbolic Species Evolved*, eds T. Schilhab, F. Stjernfelt, and T. W. Deacon (Dordrecht: Springer), 9–38. doi: 10.1007/978-94-007-2336-8\_2
- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.
- Donald, M. (1997). “Preconditions for the evolution of protolanguages,” in *The Descent of mind: Psychological Perspectives on Hominid Evolution*, eds M. C. Corballis and S. E. G. Lea (Oxford: Oxford University Press), 138–154.
- Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. New York, NY: Oxford University Press.
- Emmorey, K. (2005). Sign languages are problematic for a gestural origins theory of language evolution. *Behav. Brain Sci.* 28, 130–131. doi: 10.1017/S0140525X05270036
- Ferrari, P. F., Rozzi, S., and Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *J. Cogn. Neurosci.* 17, 212–226. doi: 10.1162/0898929053124910
- Ferry, A. L., Hespos, S. J., and Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. *Child Dev.* 81, 472–479. doi: 10.1111/j.1467-8624.2009.01408.x
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267. doi: 10.1016/S1364-6613(00)01494-7
- Fitch, W. T. (2005). “Production of vocalizations in mammals,” in *Encyclopedia of Language and Linguistics*, ed. K. Brown (New York, NY: Elsevier), 115–121.
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779
- Fitch, W. T., Hauser, M. D., and Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition* 97, 179–210; discussion 211–125. doi: 10.1016/j.cognition.2005.02.005
- Fitch, W. T., and Reby, D. (2001). The descended larynx is not uniquely human. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1669–1675. doi: 10.1098/rspb.2001.1704
- Fulkerson, A. L., and Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: evidence from 6- and 12-month-olds. *Cognition* 105, 218–228. doi: 10.1016/j.cognition.2006.09.005
- Gardner, B. T., and Gardner, R. A. (1985). Signs of Intelligence in cross-fostered chimpanzees. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 308, 159–176. doi: 10.1098/rstb.1985.0017

- Gardner, R. A., and Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science* 165, 664–672. doi: 10.1126/science.165.3894.664
- Gasser, M. (2004). “The origins of arbitrariness in language,” in *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, eds K. D. Forbus, D. Gentner, and T. Regier (Mahwah, NJ: Lawrence Erlbaum), 434–439.
- Gentilucci, M., and Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neurosci. Biobehav. Rev.* 30, 949–960. doi: 10.1016/j.neubiorev.2006.02.004
- Hauser, M. D. (2016). Challenges to the what, when, and why? *Biolinguistics* 10, 1–5.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hauser, M. D., Newport, E. L., and Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition* 78, B53–B64. doi: 10.1016/S0010-0277(00)00132-3
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M., Watumull, J., et al. (2014). The mystery of language evolution. *Front. Psychol.* 5:401. doi: 10.3389/fpsyg.2014.00401
- Hewes, G. W. (1996). “A history of the study of language origins and the gestural primacy hypothesis,” in *Handbook of Human Symbolic Evolution*, eds A. Lock and C. Peters (Oxford: Oxford University Press), 571–595.
- Heyes, C. M. (1993). Imitation, culture and cognition. *Anim. Behav.* 46, 999–1010. doi: 10.1006/anbe.1993.1281
- Hobaiter, C. L., and Byrne, R. W. (2012). “Gesture use in consortship: wild chimpanzees’ use of gesture for an ‘evolutionarily urgent’ purpose,” in *Developments in Primate Gesture Research*, eds S. Pika and K. Liebal (Amsterdam: John Benjamins), 129–146. doi: 10.1075/gs.6.08hob
- Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 88–111. doi: 10.1038/scientificamerican0960-88
- Imai, M., and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 1–13. doi: 10.1098/rstb.2013.0298
- Iverson, J. M., Hall, A. J., Nickel, L., and Wozniak, R. H. (2007). The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain Lang.* 101, 198–207. doi: 10.1016/j.bandl.2006.11.004
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198270126.001.0001
- Jürgens, U. (1976). Projections from the cortical larynx area in the squirrel monkey. *Exp. Brain Res.* 25, 401–411. doi: 10.1007/BF00241730
- Jürgens, U. (1992). “On the neurobiology of vocal communication,” in *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, eds H. Papousek, U. Jürgens, and M. Papousek (Cambridge: Cambridge University Press), 31–42.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26, 235–258. doi: 10.1016/S0149-7634(01)00068-9
- Kegl, J., Senghas, A., and Coppola, M. (1999). “Creation through contact: sign language emergence and sign language change in Nicaragua,” in *Language Creation and Language Change: Creolization, Diachrony, and Development*, ed. M. DeGraff (Cambridge, MA: MIT Press), 179–237.
- Kent, R. D. (1984). Psychobiology of speech development: coemergence of language and a movement system. *Am. J. Physiol.* 15, R888–R894. doi: 10.1152/ajpregu.1984.246.6.R888
- Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35, 379–415. doi: 10.1515/ling.1997.35.2.379
- Köhler, W. (1947). *Gestalt Psychology*, 2nd Edn. New York, NY: Liveright Publishing.
- Kuypers, H. G. (1958). Corticobular connexions to the pons and lower brainstem in man: an anatomical study. *Brain* 81, 364–388. doi: 10.1093/brain/81.3.364
- Leavens, D., Racine, T. P., and Hopkins, W. D. (2009). “The ontogeny and phylogeny of non-verbal deixis,” in *The Prehistory of Language*, eds R. P. Botha and C. Knight (New York, NY: Oxford University Press), 142–165.
- Lee, N., Mikesell, L., Joaquin, A. D. L., Mates, A. W., and Schumann, J. H. (2009). *The Interactional Instinct: The Evolution and Acquisition of Language*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195384246.001.0001
- Lee, N., and Schumann, J. H. (2005). “The interactional instinct: the evolution and acquisition of language,” in *Paper Presented at the Congress of the International Association for Applied Linguistics*, Madison, WI.
- Lemon, R. N. (2008). Descending pathways in motor control. *Annu. Rev. Neurosci.* 31, 195–218. doi: 10.1146/annurev.neuro.31.060407.125547
- Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *J. Acoust. Soc. Am.* 44, 1574–1584. doi: 10.1121/1.1911299
- Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge MA: Harvard University Press.
- Lieberman, P. (2000). *Human Language and our Reptilian Brain*. Cambridge MA: Harvard University Press.
- Lieberman, P. (2003). “Language evolution and innateness,” in *Mind, Brain, and Language Multidisciplinary Perspectives*, eds M. T. Banich and M. Mack (Mahwah, NJ: Lawrence Erlbaum), 3–21.
- Lieberman, P. (2006a). Limits on tongue deformation: Diana monkey vocalizations and the impossible vocal tract shapes proposed by Riede et al. (2005). *J. Hum. Evol.* 50, 219–221. doi: 10.1016/j.jhevol.2005.07.010
- Lieberman, P. (2006b). *Towards an Evolutionary Biology of Language*. Cambridge MA: Harvard University Press.
- Lieberman, P. (2007). The evolution of human speech. Its anatomical and neural bases. *Curr. Anthropol.* 48, 39–66. doi: 10.1086/509092
- Lieberman, P. (2016). The evolution of language and thought. *J. Anthropol. Sci.* 94, 127–146. doi: 10.4436/JASS.94029
- Lieberman, P., Crelin, E. S., and Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *Am. Anthropol.* 74, 287–307. doi: 10.1525/aa.1972.74.3.02a00020
- Lieberman, P., Laitman, J. T., Reidenberg, J. S., and Gannon, P. J. (1992). The anatomy, physiology, acoustics and perception of speech: essential elements in analysis of the evolution of human speech. *J. Hum. Evol.* 23, 447–467. doi: 10.1016/0047-2484(92)90046-C
- Livingstone, M., Srihasam, K., and Morocz, I. (2010). The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Anim. Cogn.* 13, 711–719. doi: 10.1007/s10071-010-0321-1
- Livingstone, M. S., Pettine, W. W., Srihasam, K., Moore, B., Morocz, I. A., and Lee, D. (2014). Symbol addition by monkeys provides evidence for normalized quantity coding. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6822–6827. doi: 10.1073/pnas.1404208111
- Locke, J. L. (1995). *The Child’s Path to Spoken Language*. Cambridge MA: Harvard University Press.
- Lyn, H., and Savage-Rumbaugh, E. S. (2000). Observational word learning by two bonobos: ostensive and non-ostensive contexts. *Lang. Commun.* 20, 255–273. doi: 10.1016/S0271-5309(99)00026-9
- Lynch, M. P., Oller, D. K., Steffens, M. L., and Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Dev. Psychobiol.* 28, 3–25. doi: 10.1002/dev.420280103
- MacKenzie, H., Graham, S. A., and Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Dev. Sci.* 14, 249–255. doi: 10.1111/j.1467-7687.2010.00975.x
- Maclarnon, A., and Hewitt, G. (1999). The evolution of human speech: the role of enhanced breathing control. *Am. J. Phys. Anthropol.* 109, 341–363. doi: 10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2
- Maclarnon, A., and Hewitt, G. (2004). Increased breathing control: another factor in the evolution of human language. *Evol. Anthropol. Issues News Rev.* 13, 181–197. doi: 10.1002/evan.20032
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–546. doi: 10.1017/S0140525X98001265
- MacNeilage, P. F. (2008/2010). *The Origin of Speech*. Oxford: Oxford University Press.
- MacNeilage, P. F., and Davis, B. L. (2000). On the origin of internal structure of word forms. *Science* 288, 527–531. doi: 10.1126/science.288.5465.527
- MacNeilage, P. F., and Davis, B. L. (2015). “The evolution of language,” in *The Handbook of Evolutionary Psychology*, ed. D. M. Buss (Hoboken, NJ: Wiley), 698–723.
- Magrassi, L., Aromataris, G., Cabrini, A., Annovazzi-Lodi, V., and Moro, A. (2015). Sound representation in higher language areas during language generation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1868–1873. doi: 10.1073/pnas.1418162112
- Marno, H., Farroni, T., Dos Santos, Y. V., Ekramnia, M., Nespor, M., and Mehler, J. (2015). Can you see what I am talking about? Human speech triggers

- referential expectation in four-month-old infants. *Sci. Rep.* 5:13594. doi: 10.1038/srep13594
- Marno, H., Guellai, B., Vidal, Y., Franzo, J., Nespore, M., and Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Front. Psychol.* 7:1150. doi: 10.3389/fpsyg.2016.01150
- Massaro, D. W., and Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: implications for vocabulary learning. *Front. Commun.* 2:4. doi: 10.3389/fcomm.2017.00004
- Maurer, D., Pathman, T., and Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Dev. Sci.* 9, 316–322. doi: 10.1111/j.1467-7687.2006.00495.x
- McNeill, D. (2005). *Gesture and Thought*. Chicago IL: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001
- Monaghan, P., and Christiansen, M. H. (2006). "Why form-meaning mappings are not entirely arbitrary in language," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, ed. R. Sun (Hillsdale, NJ: Lawrence Erlbaum), 1838–1843.
- Monaghan, P., Christiansen, M. H., and Fitneva, S. A. (2011). The arbitrariness of the sign: learning advantages from the structure of the vocabulary. *J. Exp. Psychol. Gen.* 140, 325–347. doi: 10.1037/a0022924
- Monaghan, P., Mattock, K., and Walker, P. (2012). The role of sound symbolism in language learning. *J. Exp. Psychol. Mem. Cogn.* 38, 1152–1164. doi: 10.1037/a0027747
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., and Kirby, S. (2014). How arbitrary is language? *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130299. doi: 10.1098/rstb.2013.0299
- Morton, E. S. (1994). "Sound symbolism and its role in non-human vertebrate communication," in *Sound Symbolism*, eds L. Hinton, J. Nichols, and J. J. Ohala (Cambridge: Cambridge University Press), 348–365.
- Negus, V. E. (1949). *The Comparative Anatomy and Physiology of the Larynx*. New York, NY: Grune and Stratton.
- Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: evolution of face flattening and speech. *J. Hum. Evol.* 51, 244–254. doi: 10.1016/j.jhevol.2006.03.005
- Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005
- Ohala, J. J. (1994). "The frequency code underlies the sound-symbolic use of voice pitch," in *Sound Symbolism*, eds L. Hinton, J. Nichols, and J. J. Ohala (Cambridge: Cambridge University Press), 325–365.
- Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah NJ: Lawrence Erlbaum. doi: 10.4324/9781410602565
- Oller, D. K., and Griebel, U. (2014). On quantitative comparative research in communication and language evolution. *Biol. Theory* 9, 296–308. doi: 10.1007/s13752-014-0186-7
- Oller, D. K., and Lynch, M. P. (1992). "Infant vocalizations and innovations in infraphonology: toward a broader theory of development and disorders," in *Phonological Development: Models, Research, Implications*, eds C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Timonium, MD: York Press), 509–536.
- Oudeyer, P.-Y., and Kaplan, F. (2006). Discovering communication. *Conn. Sci.* 18, 189–206. doi: 10.1080/09540090600768567
- Pedersen, J. (2012). "The symbolic mind: apes, symbols, and the evolution of language," in *Paper Presented at the Graduate Theses and Dissertations* 12430, (Ames, IA: Iowa State University).
- Peirce, C. S. (1998). "What is a sign?," in *The Essential Peirce: Selected Philosophical Writings*, Vol. 2, ed. P. E. Project (Bloomington IN: Indiana University Press), 1893–1913.
- Penn, D. C., and Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 731–744. doi: 10.1098/rstb.2006.2023
- Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *R. Soc. Open Sci.* 2:150152. doi: 10.1098/rsos.150152
- Perruchet, P., Peereman, R., and Tyler, M. D. (2006). Do we need algebraic-like computations? A reply to Bonatti, Pena, Nespore, and Mehler (2006). *J. Exp. Psychol. Gen.* 135, 322–326. doi: 10.1037/0096-3445.135.2.322
- Perruchet, P., and Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon. Bull. Rev.* 12, 307–313. doi: 10.3758/BF03196377
- Perruchet, P., Tyler, M. D., Galland, N., and Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *J. Exp. Psychol. Gen.* 133, 573–583. doi: 10.1037/0096-3445.133.4.573
- Pika, S., Liebal, K., Call, J., and Tomasello, M. (2005). The gestural communication of apes. *Gesture* 5, 41–56. doi: 10.1075/gest.5.1-2.05pik
- Pinker, S. (1994). *The Language Instinct*. New York, NY: William Morrow.
- Pinker, S., and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236. doi: 10.1016/j.cognition.2004.08.004
- Pleyer, M. (2017). Protolanguage and mechanisms of meaning construal in interaction. *Lang. Sci.* 63, 69–90. doi: 10.1016/j.langsci.2017.01.003
- Pleyer, M., and Winters, J. (2015). Integrating cognitive linguistics and language evolution research. *Theoria Hist. Sci.* 11, 19–43. doi: 10.12775/ths-2014-002
- Pollick, A. S., and de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8184–8189. doi: 10.1073/pnas.0702624104
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., and Fischer, J. (2015). Vervets revisited: a quantitative analysis of alarm call structure and context specificity. *Sci. Rep.* 5:13220. doi: 10.1038/srep13220
- Provine, R. R. (2005). Walkie-talkie evolution: bipedalism and vocal production. *Behav. Brain Sci.* 27, 520–521. doi: 10.1017/S0140525X04410115
- Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi: 10.1016/S0166-2236(98)01260-0
- Sandler, W. (2012). *Dedicated Gestures and the Emergence of Sign Language*. 265–307. Available at: [http://sandlerlab.haifa.ac.il/html/html\\_eng/Dedicated\\_Gestures.pdf](http://sandlerlab.haifa.ac.il/html/html_eng/Dedicated_Gestures.pdf)
- Sandler, W. (2013). Vive la différence: sign language and spoken language in language evolution. *Lang. Cogn.* 5, 189–203. doi: 10.1515/langcog-2013-0013
- Sandler, W. (2017). The challenge of sign language phonology. *Annu. Rev. Linguist.* 7, 43–63. doi: 10.3758/s13428-015-0579-y
- Sandler, W., Aronoff, M., Padden, C., and Meir, I. (2014). "Language emergence Al-sayyid bedouin sign language," in *The Cambridge Handbook of Linguistic Anthropology*, eds N. J. Enfield, P. Kockelman, and J. Sidnell (Cambridge: Cambridge University Press), 246–278.
- Sapir, E. (1929). A study in phonetic symbolism. *J. Exp. Psychol.* 12, 225–239. doi: 10.1037/h0070931
- Sasaki, S., Isa, T., Pettersson, L.-G., Alstermark, B., Naito, K., Yoshimura, K., et al. (2004). Dexterous finger movements in primate without monosynaptic corticomotoneuronal excitation. *J. Neurophysiol.* 92, 3142–3147. doi: 10.1152/jn.00342.2004
- Saussure, F. D. (1916/1966). *Course in General Linguistics*. New York, NY: McGraw-Hill.
- Savage-Rumbaugh, E. S., and Fields, W. (2000). Linguistic, cultural and cognitive capacities of bonobos (*Pan paniscus*). *Cult. Psychol.* 6, 131–153. doi: 10.1177/1354067X00062003
- Savage-Rumbaugh, E. S., McDonald, K., Sevcik, R. A., Hopkins, W. D., and Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *J. Exp. Psychol. Gen.* 115, 211–235. doi: 10.1037/0096-3445.115.3.211
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., and Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monogr. Soc. Res. Child Dev.* 58, 1–221. doi: 10.2307/1166068
- Savage-Rumbaugh, E. S., and Rumbaugh, D. M. (1978). Symbolization, language, and chimpanzees: a theoretical reevaluation based on initial language acquisition processes in four young *Pan troglodytes*. *Brain Lang.* 6, 265–300. doi: 10.1016/0093-934X(78)90063-9
- Savage-Rumbaugh, S., Shanker, S. G., and Taylor, T. J. (2001). *Apes, Language, and the Human Mind*. New York, NY: Oxford University Press.
- Savage-Rumbaugh, S. E., and Lewin, R. (1996). *Kanzi: The ape at the Brink of the Human Mind*. New York, NY: Wiley.
- Senghas, A. (1995). *Children's Contribution to the Birth of Nicaraguan Sign Language*. Doctoral dissertation, Massachusetts Institute of Technology, Boston, MA.
- Sereno, M. I. (2014). Origin of symbol-using systems: speech, but not sign, without the semantic urge. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130303. doi: 10.1098/rstb.2013.0303

- Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094. doi: 10.1016/S0003-3472(80)80097-2
- Simonyan, K. (2014). The laryngeal motor cortex: its organization and connectivity. *Curr. Opin. Neurobiol.* 28, 15–21. doi: 10.1016/j.conb.2014.05.006
- Simonyan, K., and Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist* 17, 197–208. doi: 10.1177/1073858410386727
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174
- Sonnweber, R., Ravignani, A., and Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Anim. Cogn.* 18, 733–745. doi: 10.1007/s10071-015-0840-x
- Squire, L. R., and Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13515–13522. doi: 10.1073/pnas.93.24.13515
- Stjernfelt, F. (2012). “The evolution of semiotic self-control,” in *The Symbolic Species Evolved*, eds T. Schilhab, F. Stjernfelt, and T. W. Deacon (Dordrecht: Springer), 39–63. doi: 10.1007/978-94-007-2336-8\_3
- Studdert-Kennedy, M. (1998). “Introduction: the emergence of phonology,” in *Approaches to the Evolution of Language: Social and Cognitive Bases*, eds J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Cambridge: Cambridge University Press).
- Sutton, D., Larson, C., Taylor, E. M., and Lindeman, R. C. (1973). Vocalization in rhesus monkeys: conditionability. *Brain Res.* 52, 225–231. doi: 10.1016/0006-8993(73)90660-4
- Thom, S. A., Hoit, J. D., Hixon, T. J., and Smith, A. E. (2006). Velopharyngeal function during vocalization in infants. *Cleft Palate Craniofac. J.* 43, 539–546. doi: 10.1597/05-113
- Thompson, R. L., Vinson, D. P., Woll, B., and Vigliocco, G. (2012). The road to language learning is iconic: evidence from British Sign Language. *Psychol. Sci.* 23, 1443–1448. doi: 10.1177/0956797612459763
- Tomasello, M. (1990). “Cultural transmission in the tool use and communicatory signaling of chimpanzees?,” in *Language and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives*, eds S. T. Parker and K. R. Gibson (New York, NY: Cambridge University Press), 274–311.
- Tomasello, M. (1996). “Do apes ape?,” in *Social Learning in Animals: the Roots of Culture*, eds C. M. Heyes and B. G. Galef (San Diego, CA: Academic Press), 319–346. doi: 10.1016/B978-012273965-1/50016-9
- Tomasello, M., and Zuberbühler, K. (2002). “Primate vocal and gestural communication,” in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds M. Bekoff, C. Allen, and G. M. Burghardt (Cambridge, MA: MIT Press), 293–299.
- Vale, G. L., Davis, S. J., Lambeth, S. P., Schapiro, S. J., and Whiten, A. (2017). Acquisition of a socially learned tool use sequence in chimpanzees: implications for cumulative culture. *Evol. Hum. Behav.* 38, 635–644. doi: 10.1016/j.evolhumbehav.2017.04.007
- Vale, G. L., Flynn, E. G., Pender, L., Price, E., Whiten, A., Lambeth, S. P., et al. (2016). Robust retention and transfer of tool construction techniques in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 130, 24–35. doi: 10.1037/a0040000
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., et al. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences. *Psychol. Sci.* 21, 21–25. doi: 10.1177/0956797609354734
- Wechsler, D. (1997). *WAIS-III, Wechsler adult Intelligence Scale*. San Antonio TX: Psychological Corporation.
- Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 2417–2428. doi: 10.1098/rstb.2009.0069
- Williams, P., Beer, R., and Gasser, M. (2008). “Evolving referential communication in embodied dynamical agents,” in *Paper Presented at the International Conference on the Simulation and Synthesis of Living Systems*, (Cambridge, MA: MIT Press).
- Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., and Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science* 300, 1578–1581. doi: 10.1126/science.1084324

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Boucher, Gilbert and Rossier-Bisailon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.