



Vocal Development as a Guide to Modeling the Evolution of Language

D. Kimbrough Oller,^a Ulrike Griebel,^a Anne S. Warlaumont^b

^a*University of Memphis and Konrad Lorenz Institute for Evolution and Cognition Research*
^b*Cognitive and Information Sciences, University of California, Merced*

Received 28 March 2014; received in revised form 16 June 2015; accepted 4 July 2015

Abstract

Modeling of evolution and development of language has principally utilized mature units of spoken language, phonemes and words, as both targets and inputs. This approach cannot address the earliest phases of development because young infants are unable to produce such language features. We argue that units of early vocal development—protophones and their primitive illocutionary/perlocutionary forces—should be targeted in evolutionary modeling because they suggest likely units of hominin vocalization/communication shortly after the split from the chimpanzee/bonobo lineage, and because early development of spontaneous vocal capability is a logically necessary step toward vocal language, a root capability without which other crucial steps toward vocal language capability are impossible. Modeling of language evolution/development must account for dynamic change in early communicative units of form/function across time. We argue for interactive contributions of sender/infants and receiver/caregivers in a feedback loop involving both development and evolution and propose to begin computational modeling at the hominin break from the primate communicative background.

Keywords: Vocal development; Language evolution; Computational modeling; Spontaneous vocalization; Parent-infant interaction; Illocution; Perlocution

1. Spontaneous infant vocalization and foundations of language

Among the diverse theories on origins of language, the focus has been on high-level language capabilities (phonemes, words, sentences...) rather than on prelinguistic vocal communication as a logical precursor, a foundation without which other crucial capabilities for language cannot be built. Our goals here are to argue for an evo-devo approach

Correspondence should be sent to D. K. Oller, University of Memphis, School of Communication Sciences and Disorders, 4055 N. Park Loop, Memphis, TN 38152. E-mail: koller@memphis.edu

to language evolution using human development as a guide and to propose a computational modeling agenda for language evolution, focusing on the emergence of *Spontaneous Vocalization*. First we provide reasoning behind this proposal and then offer a sketch of proposed computational modeling studies, focused initially on emergence of *Spontaneous Vocalization* with a view toward subsequent modeling of the three additional capabilities depicted in Fig. 1 (see explanations below).

In the first days of life, human infants produce seemingly *Spontaneous Vocalizations*, “protophones” (Oller, 2000), often in apparent comfort and often in the absence of social interaction. Protophones are not vegetative sounds such as coughs or burps, nor are they

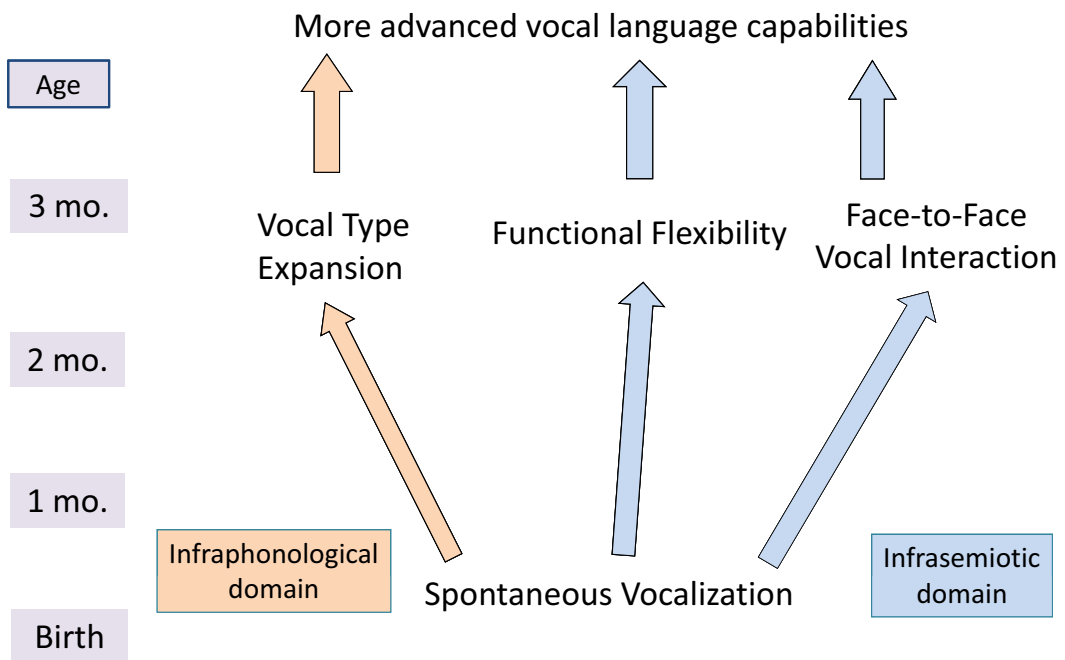


Fig. 1. Schematic fragment of communicative natural logic: Tree of vocal development/evolution.

Notes. The infrastructural tree shows vocal capabilities in order of emergence in human development from bottom up, infraphonological/signal capabilities on the left, infrasemiotic/function capabilities on the right. At birth, *Spontaneous Vocalization* in modern humans includes the endogenous tendency to produce phonation, which we propose occurred in the human line at the beginning of differentiation from non-human primate vocal communication. On the basis of *Spontaneous Vocalization*, additional capabilities emerge by 3 months: (a) *Vocal Type Expansion* (resulting in protophones such as vowel-like sounds, squeals, and growls) appears to result from self-organization through phonatory exploration, (b) *Functional Flexibility*, where each protophone is used with a full range of illocutionary/perlocutionary functions, expressing infant positivity, neutrality or negativity, and eliciting corresponding caregiver responses (Oller et al., 2013), and (c) *Face-to-Face Vocal Interaction*, where infants appear to bond with parents in conversation. The figure illustrates a hypothesis about how language foundations *do* and presumably *must* develop and evolve. Such capabilities have been called “preadaptations” for language (Christiansen & Kirby, 2003). See Supplementary Material, section SM2 for further explanation and for description of subsequent stages leading to spoken language.

effort grunts; we interpret them instead as endogenously generated precursors to speech. By 3 months, the protophones differentiate into at least three types based on phonation¹ and pitch. These include (a) vowel-like sounds produced in infants' mid-pitch range with normal phonation, the kind of phonation that predominates in speech; (b) squeals with high pitch, often in falsetto phonation; and (c) growls with either low pitch, creaky phonation, or raucous dysphonation (Holmgren, Lindblom, Aurelius, Jalling, & Zetterstrom, 1986; Oller, 1980; Stark, Rose, & McLagen, 1975) (audio-video examples at babyvoc.org, IVICT). Also by 3 months, all the protophones are accompanied by positive, neutral, and negative facial affect, a pattern suggesting "functional flexibility" (Oller et al., 2013), where *Spontaneous Vocalizations* on different occasions transmit different emotions and illocutionary forces and elicit different caregiver responses, that is, perlocutionary effects (see Supplementary Material, section SM1 for illocution/perlocution definitions, concepts adapted from Austin, 1962). In addition, by 3 months, the protophones are produced both in solitary activity and in elaborate face-to-face vocal interactions, often lasting minutes, and appearing to constitute protoconversational bonding events between caregivers and infants (Stern, Jaffe, Beebe, & Bennett, 1975; Tronick & Cohn, 1989).

Cry and laughter, sounds thought more similar to "calls" in non-human primates (Owren, Amoss, & Rendall, 2011), show strong functional bias, cry expressing negative and laugh positive emotion. Functional flexibility of the protophones has been interpreted as a foundation for speech, since in speech all syllables, words and sentences can be used with differing functions on different occasions. Spontaneous protophones occur far more frequently than cry/laughter (Nathani, Ertmer, & Stark, 2006) even in the first months and are also observed in preterm infants at 32 and 36 weeks gestational age (Oller et al., 2014). The pattern of high spontaneous volubility² from very early appears to be without precedent in other apes, suggesting selection pressure in the hominin line on spontaneous, exploratory vocalization. Other apes have also never been reported to produce any vocalization type with affect ranging from positive, to neutral, to negative, nor to engage in sustained face-to-face vocal interactions in the absence of aggression.

2. Infrastructural framework

The differentiating properties of the early protophones are predominantly phonatory,¹ rather than articulatory or intonational (Buder, Chorna, Oller, & Robinson, 2008). Such sounds are infrastructural in the sense that controlled phonation is required as a starting point in vocal language, with the vast majority of syllables in natural languages requiring normally phonated nuclei.

As illustrated in Fig. 1, *Spontaneous Vocalization* forms naturally logical infrastructure for three additional capabilities: (a) *Vocal Type Expansion* yields the protophones, seemingly self-organized through vocal exploration; (b) these vocal types are characterized by *Functional Flexibility*, expressing varying emotional states, and eliciting varying caregiving responses on different occasions of use; and (c) these types provide anchors for elaborate *Face-to-Face Vocal Interaction* with caregivers. Each of these in turn serves as

infrastructure for additional capabilities required for vocal language, such as the capability to flexibly direct sounds to caregivers, to use vocalization to support joint attention, and so on (as explicated in SM2). The human infant goes through these steps of capability-building, starting at the bottom of the tree and cannot begin to speak meaningfully in words or sentences until a number of additional capabilities beyond those in Fig. 1 (see SM2) have also been developed. Thus, all aspects of speech communication appear to depend upon being able to vocalize at will, at any time, and for any purpose (Oller, 2000).

3. Selection for spontaneous vocalization

The early protophones (the infraphonological material used in transmission) do not transmit meanings, as words can, but rather reveal fitness or express states (the infrasemiotic material transmitted *by* the protophones at the beginning of life). We make no assumption that very young infants intend to communicate with protophones. On the contrary, the first *Spontaneous Vocalizations* show no indication of being directed to any receiver. So how could undirected vocalizations function to the selective benefit of infants?

Our proposal assumes that ancient infant hominins were, more than other infant primates, subject to caregiver selection pressures and that differences among infants in vocalization helped caregivers determine how much to invest in individual infants. Relative altriciality³ of the hominin infant in comparison to other apes (Bogin, 1990) may have increased selection pressure on fitness signals from infants (Locke & Bogin, 2006), since a longer period of infant helplessness than in related species would have favored longer-term parental investment (Locke, 2006; Oller & Griebel, 2008). High rate and diversity of protophone production appears to constitute a true fitness signal in much the same way that playful behavior in general suggests well-being (Lafreniere, 2011) and is “rightly regarded as a useful index of the physical and psychological well-being of the young primate” (Mason, 1965, p. 530). Furthermore, many observers have suggested that non-cry vocalization in infancy can be interpreted by caregivers as a sign of comfort and presumable well-being (Shimada, 2012; Stark et al., 1975). Thus, the initial functions of *Spontaneous Vocalization* appear to be signaling of fitness and well-being, with caregivers responding nurturantly (a perlocutionary effect, see SM1). Additional functions emerge in modern human infants from *Spontaneous Vocalization*, and as development proceeds, infant capability for intentional action grows and vocalization comes to be used in new ways (Fig. 1, see SM1-SM2 for elaborations).

The proposal also assumes hominin receivers had/have broad reception/interpretation capabilities allowing them to provide differentiated feedback to signals. This assumption is consistent with results in animal communication indicating primates are far more flexible in recognition/reaction to signals than in signal production or functional usage (Owren et al., 2011; Seyfarth & Cheney, 1997). Flexible responsivity of receivers is a necessary

condition for their being able to apply selection pressure on volubility² and flexibility of infant vocal tendencies.

Vocal fitness signals could have initially become a focus for hominin parental investment if (a) preadaptation for vocal control had already evolved, making *Spontaneous Vocalization* more frequent, with accompanying increase in the tendency for infants to explore vocalization, (b) increased size of hominin social groups (Dunbar, 1996) had reduced the presumable pan-ape priority on silence to avoid detection by predators, (c) larger groups had raised priority on vocal interaction to replace grooming as a mechanism of alliance formation and reconciliation (Dunbar, 1996), or (d) any combination of these. Physiological foundations for *Spontaneous Vocalizations* without social purpose have been studied in infant rats, with results suggesting phylogenetically early foundations for *Spontaneous Vocalizations* in human infants, where social purpose (from the infant standpoint) is largely absent at first but develops with time (Blumberg & Alberts, 1990; Blumberg, Gravato Marques, & Iida, 2013).

Our proposed computational modeling will focus first on the seemingly simple matter of how selection pressures might engender increase in the tendency of an infant (or computational agent) to produce *Spontaneous Vocalization*. In our approach, selection pressures at each stage of vocal development and evolution are supplied by Receivers/caregivers, who respond to stage-appropriate Sender/infant vocalizations with systematic nurturance (Fig. 2A). The approach is consistent with the idea that *evolution* of increased *Spontaneous Vocalization* began with *developmental* steps in individual infants under the selective pressure of their own caregivers. Indeed, during modern human development, infant vocal capabilities emerge partly in response to social interaction (and increasingly so as the infant matures), where caregivers react to vocal capabilities of infants in accord with a scaffolding principle requiring parental discernment (Bruner, 1985) and “intuitive parenting” to reinforce vocal exploration and learning (Papoušek & Papoušek, 1987). Both endogenous inclinations of infants to explore the vocal space and interactive feedback from caregivers thus foster growth in vocal capability (Goldstein & Schwade, 2008; Goldstein, Schwade, & Bornstein, 2009).

If our assumptions about selection pressures for *Spontaneous Vocalization* are valid, it follows that hominin parents would *also* have been selected to become aware of the fitness reflected in infant vocalizations and capable of responding to those indicators with selective care and reinforcement of vocalizations. Thus, we reason, each aspect of the feedback loop depicted in Fig. 2A (see SM4 for details and citations on empirical support for features of the loop) would have been under selection pressure, creating a dynamic system that would have fostered growth through the steps implied by Fig. 1 and Fig. SM at a rate that computational modeling should ultimately help to estimate.

The vast majority of evolutionary computational modeling has focused on elements of mature language-like behavior (syllables, words, etc., see SM2–SM3 for discussion and references). But recently, computational modeling of early vocal development has also become active (Oudeyer & Kaplan, 2006; Moulin-Frier, Nguyen, & Oudeyer, 2014; Warlaumont, Westermann, Buder, & Oller, 2013; Warlaumont, Richards, Gilkerson, & Oller, 2014, and see SM3). Such work both instantiates theoretical constructs regarding language foundations and tests interactions among mechanisms such as infant vocal

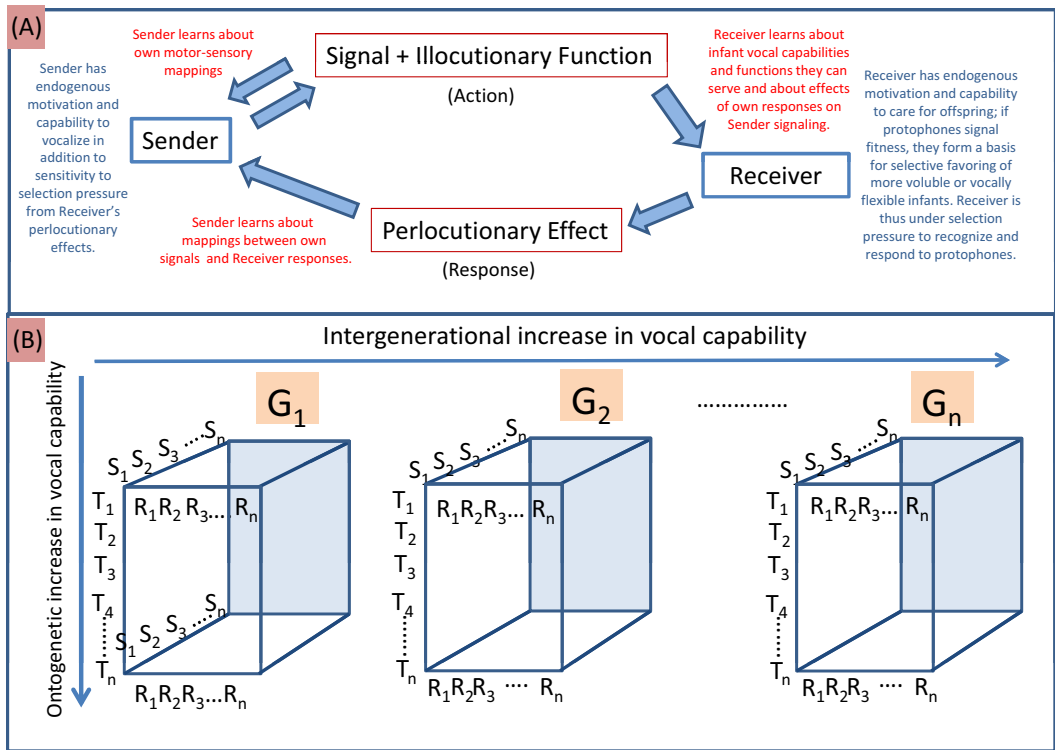


Fig. 2. Computational modeling of interactive feedback in communicative development/evolution.

Notes. (A) In our proposed computational modeling of communication, a Signal produced (and sensed) by a Sender/infant, transmits an Illocutionary Function recognized by a Receiver/caregiver, whose response is a Perlocutionary Effect, which upon reception by the Sender serves as a selection force on the Sender/infant's capability/inclination to produce the Signal/Function (terms adapted from Austin [1962], see SM1). Learning by both Sender and Receiver can occur at various points in the computationally modeled feedback loop, which can be tested for outcomes and initial settings on (1) infant volubility, (2) infant self-monitoring in signal production, (3) infant learning sensitivity to parental responses, (4) parent responsivity to infant signals, (5) parent adjustment of responsivity based on infant capabilities previously exhibited, and (6) parent self-monitoring of the effects of their own responsivity on infant signaling. Correlations between these outcomes/settings and reproductive success can also be tested and assessed in reference to the extensive literature in vocal and interaction development (see section SM4). (B) Feedback will operate in the computational models in both intra- and inter-generational selection, in the former case selecting for vocal *development* in individual Senders (S_1 – S_n) under the influence of Receivers (R_1 – R_n) *within* generations, and in the latter case selecting Senders of the species in vocal *evolution* across generations. Senders of the first generation (G_1) will be differentially selected to reproduce in G_2 , such that progeny of more vocally capable Senders from G_1 will be overrepresented in G_2 . Additionally, more vocally capable Senders from G_1 will also be overrepresented as influential Receivers in G_2 , since they will be more likely to have children in G_2 than less vocally capable Senders from G_1 .

production tendencies and caregiver responsivity within and across generations, yielding possible empirical evaluation of relative roles for and interactions between evolution and development (learning) in language emergence.

Based in part on these recent efforts, our proposed agenda for further research: (a) starts at the break from the chimpanzee/bonobo background, the tree's trunk, Fig. 1; and (b) addresses interactive feedback (Fig. 2A), where Senders produce signals/functions, Receivers provide selection pressure on production capabilities, selection pressure also applies to Receivers' detection and responses to correlations between vocalizations and infant fitness, and (c) the feedback loop operates at varying time scales from short periods of learning by individual organisms, to cross-generational cultural evolution, to deep time evolution of innate predispositions (Fig. 2B).⁴

4. Example of proposed approach

Fig. 2 illustrates components to include in our proposed computational modeling agenda to address spontaneous protophone evolution, testing scenarios, and mechanisms with the potential to evolve Senders/infants that spontaneously produce high numbers of protophones (high volubility²) correlated with their fitness, and Receivers/parents that respond nurturantly to infants with higher volubility. At the beginning of each modeling run, Senders/infants will have an endogenous tendency to vocalize randomly at some minimal level, the ability to learn, and interest in learning about sensory and social consequences of their motor actions. Receivers/parents will have an endogenous interest in providing greater nurturance to infants deemed more fit because of their high volubility. Receivers will also have an endogenous ability to learn from previous experiences with infant vocalizations, in particular to learn what their infant's current repertoire and rate of volubility is and to adjust their responses to reinforce infant vocalizations that demonstrate progress in increasing repertoire size and/or volubility. Additionally, features of Receiver responsiveness, such as the tendency to invest nurturance in their infants, base rate of attention to infant volubility level, and learning abilities will evolve across generations.

There are at least two relevant modeling timescales for adaptation, shown in Fig. 2B: On an ontogenetic timescale, both infants and caregivers adapt through learning (red lettering, Fig. 2A), while on a phylogenetic timescale, adaptation occurs through evolution of Sender/infant and Receiver/parent capabilities for the relevant actions and for learning them (blue lettering). Learning can occur in several ways, as indicated in the figure, all of which can be separately manipulated parametrically during modeling, with settings guided by empirical results from human development (see SM4).

Sensible initial modeling goals are to determine parameters that yield increased *Spontaneous Vocalization* (higher protophone volubility) in infants, along with higher responsiveness in parents, and to assess time frames of such increases for both ontogeny and phylogeny. As indicated in Fig. 2B, each generation of Receivers/parents should be selected on the basis of volubility and volubility learning from the prior generation, in keeping with our hypothesis that volubility should correlate with fitness. Modeling results can be judged in terms of plausibility against existing data on volubility in human infants and can be cast in light of current speculations about time frames of hominin

vocal evolution. Such modeling can operate at varying levels, taking into account genetics as well as factors such as vocal tract anatomy and neurophysiology, where knowledge about epigenetic development of these factors can be expected to improve estimates.

After modeling evaluations on infant volubility (*Spontaneous Vocalization*), we recommend assessing additional capabilities in Fig. 1: *Vocal Type Expansion*, *Functional Flexibility*, and *Face-to-Face Vocal Interaction*. These tests will allow assessment of learning that leads to progressively more varied infant vocalizations (*Vocal Type Expansion*) via vocal exploration and self-organized learning, to infants gaining curiosity and knowledge about their own motor systems, and to parent learning of how best to reinforce more varied vocalizations. The acoustic content of the sounds can be modeled to mimic real infant phonatory categories and their patterns of production. Similarly, modeling can address relations between infant vocalizations and social, emotional, or physical states or contexts (*Functional Flexibility* as well as form-function mappings). Such learning can be expected to lead to infant vocalizations showing both flexible mappings as are required for arbitrary form/meaning associations in language (parents can be modeled to recognize this infant capability as a fitness signal), and acquisition of statistically reliable form-function mappings, where infant and parent may converge, yielding learning of favored mappings. Finally we suggest assessing the degree to which the above modeling approaches result in growth of *Face-to-Face Vocal Interaction*, increasing and refining parental investment, heightening *Vocal Type Expansion*, and diversifying functions served by infant vocalizations.

Additional phases of intra-generational selection pressures can also be introduced to make the modeling more realistic. Beyond infancy, peer-to-peer alliance formation (Senders become Receivers applying selection force on other Senders) can be modeled. Adult Senders can also be designated male and female and exert mutual sexual selection pressure on adult Receivers based on vocal sophistication. In these studies, interactions among various Sender-Receiver pairings will span the life-time and be assessed in terms of growth in volubility and vocal sophistication and the role of social relations in that growth.

As indicated above, computational modeling work has already begun to address some subcomponents of the approach we propose here. One body of work has modeled intrinsic motivation in infants to assess motor-sensory mappings, to monitor regions of motor/acoustic space where high learning progress is being made and to explore those regions preferentially (e.g., Moulin-Frier & Oudeyer, 2012; Moulin-Frier et al., 2014). These models show vocalizations increasing in sophistication over time. Complementary work has modeled infant progression toward increased production of more advanced vocalization types through contingent reinforcement from social or internal sources (Warlaumont, 2013; Warlaumont et al., 2013, 2014).

Thus progress is being made toward modeling infant vocal learning. More work on the possible role of vocal learning in language evolution is needed. Little attention has been paid to modeling learning by caregivers who provide the selection force on prelinguistic vocalization capabilities, nor on targeting the endogenous interests and learning abilities

of both infants and caregivers and their interaction. Our approach has the advantages of focusing modeling on the beginning of the break by the human line from that of the rest of the great apes and of allowing a more comprehensive account of adaptation on the part of both infant and parent at both phylogenetic and ontogenetic timescales.

Acknowledgments

This research was supported by NIDCD R01 DC011027 and by the Plough Foundation.

Notes

1. Phonation and phonatory sounds are produced by vibration of the vocal cords with no necessary articulation by lips, tongue, or jaw.
2. Volubility is rate of vocalization. Proto-phoneme volubility is the number of proto-phonemes produced per unit time.
3. Altricial animals (e.g., humans) are born relatively helpless, in contrast with precocial animals (e.g., other apes), who tend to locomote, climb, and forage earlier in life.
4. In the future, it will be important to include counter-pressure(s) *against Spontaneous Vocalization* in modeling the split between hominins and other primates. An obvious possibility is that vocalization may alert predators, yielding selection pressure for silence. Large group sizes in hominins (Dunbar, 1996) may have mitigated this counter-pressure.

References

- Austin, J. L. (1962). *How to do things with words*. London: Oxford University Press.
- Blumberg, M. S., & Alberts, J. R. (1990). Ultrasonic vocalizations by rat pups in the cold: An acoustic by-product of laryngeal braking? *Behavioral Neuroscience*, *104*(5), 808–817.
- Blumberg, M. S., Gravato Marques, H., & Iida, F. (2013). Twitching in sensorimotor development from sleeping rats to robots. *Current Biology*, *23*(12), 532–537.
- Bogin, B. (1990). The evolution of human childhood. *BioScience*, *40*, 16–25.
- Bruner, J. (1985). Vygotsky: A historical and conceptual perspective. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 21–34). Cambridge: Cambridge University Press.
- Buder, E. H., Chorna, L., Oller, D. K., & Robinson, R. (2008). Vibratory regime classification of infant phonation. *Journal of Voice*, *22*, 553–564.
- Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, *7*(7), 300–307.
- Dunbar, R. I. M. (1996). *Gossiping, grooming and the evolution of language*. Cambridge, MA: Harvard University Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–522.

- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from adults. *Child Development, 80*(3), 636–644.
- Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterstrom, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 51–63). New York: Stockton Press.
- Lafreniere, P. J. (2011). Evolutionary functions of social play: Life histories, sex differences, and emotion regulation. *American Journal of Play, 3*(4), 464–488.
- Locke, J. L. (2006). Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. *Human Nature, 17*, 155–168.
- Locke, J., & Bogin, B. (2006). Language and life history: A new perspective on the evolution and development of linguistic communication. *Behavioral & Brain Sciences, 29*, 259–325.
- Mason, W. A. (1965). The social development of monkeys and apes. In I. DeVore (Ed.), *Primate behavior: Field studies of monkeys and apes* (pp. 514–544). New York: Holt, Rinehart & Winston.
- Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P.-Y. (2014). Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Psychology, 4*, 1–20. doi:10.3389/fpsyg.2013.01006.
- Moulin-Frier, C., & Oudeyer, P.-Y. (2012). Curiosity-driven phonetic learning. Proceedings of IEEE, Development and Learning and Epigenetic Robotics, (ICDL), 2012 IEEE International Conference, pp. 1–8.
- Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics and Phonetics, 20*(5), 351–367.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology, Vol 1: Production* (pp. 93–112). New York: Academic Press.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences, 110*, 6318–6632.
- Oller, D. K., & Griebel, U. (2008). Complexity and flexibility in infant vocal development and the earliest steps in the evolution of language. In D. K. Oller & U. Griebel (Eds.), *Evolution of communicative flexibility: Complexity, creativity and adaptability in human and animal communication* (pp. 141–168). Cambridge, MA: MIT Press.
- Oller, D. K., Vohr, B. R., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Warlaumont, A. S., Griebel, U. & Buder, E. H. (2014). Infant vocalization and birdsong: An infrastructural view. Paper presented at the Society for Neuroscience satellite meeting: Birdsong4: Rhythms and Clues from Neurons to Behavior, Washington, DC.
- Oudeyer, P.-Y., & Kaplan, F. (2006). Discovering communication. *Connection Science, 18*(2), 189–206.
- 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.
- Papoušek, H., & Papoušek, M. (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 669–720). New York: Wiley.
- Seyfarth, R. M., & Cheney, D. L. (1997). Behavioral mechanisms underlying vocal communication in non-human primates. *Animal Learning & Behavior, 25*(3), 249–267.
- Shimada, Y. M. (2012). Infant vocalization when alone: Possibility of early sound playing. *International Journal of Behavioral Development, 36*(6), 407–412.
- Stark, R. E., Rose, S. N., & McLagen, M. (1975). Features of infant sounds: The first eight weeks of life. *Journal of Child Language, 2*, 205–222.
- Stern, D. N., Jaffe, J., Beebe, B., & Bennett, S. L. (1975). Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. *Annals of the New York Academy of Sciences, 263*, 89–100.

- Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, 60, 85–92.
- Warlaumont, A. S. (2013). Saliency-based reinforcement of a spiking neural network leads to increased syllable production. In *Proceedings of the 2013 IEEE international conference on development and learning and epigenetic robotics (ICDL)*. Osaka, pp. 1–7. doi:10.1109/DevLrn.2013.6652547
- Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, 25(7), 1314–1324.
- Warlaumont, A. S., Westermann, G., Buder, E. H., & Oller, D. K. (2013). Prespeech motor learning in a neural network using reinforcement. *Neural Networks*, 38, 64–95.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article :

Supplementary Material SM1. Intrasemiotic content in signaling: Perlocution, illocution, and meaning.

Supplementary Material SM2. The infrastructural natural logic of vocal language.

Supplementary Material SM3. Prior computational modeling of the evolution of language and early human vocal development.

Supplementary Material SM4. Features of interaction depicted in Fig. 2 of the main text.