2020

Endogenous and social factors influencing infant vocalizations as fitness signals

Helen L. Long

ENDOGENOUS AND SOCIAL FACTORS INFLUENCING INFANT VOCALIZATIONS AS FITNESS SIGNALS

by

Helen L. Long

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

University of Memphis

December 2020

Dedication

To my advisors, including Dr. D. Kimbrough Oller, who mercilessly pushed me when I needed it the most: I am unbelievably grateful for your guidance. To my family, for your unconditional support and encouragement throughout my life and education: you are the reason I made it this far. To my husband Nathan, for your love, humor, and unwavering support: I am the luckiest woman in the world to have had you with me since day one—Publication cupcakes keep me going! And in memory of the astounding Dr. Lisa Scott, who inspired me to pursue a PhD from the beginning. *"This is the first step. So send it already!"*

Acknowledgements

This work was supported by the Plough Foundation and grants from the National Institute of Deafness and Communication Disorders of the National Institutes of Health, R01DC015108, R01DC011027, and R01DC006099, awarded to D. Kimbrough Oller. Scholarly travel funding was also provided by the Institute of Intelligent Systems Student Organization and the Graduate Student Association of the University of Memphis.

Preface

This dissertation was written using the journal-ready article format. Chapter 2 of this dissertation was published as a journal article in 2019 in *Frontiers of Psychology*. Its authors are Helen L. Long, D. Kimbrough Oller, and Dale D. Bowman. Chapter 3 was published in 2020 in *PLoS ONE*. Its authors are Helen L. Long, Dale D. Bowman, Hyunjoo Yoo, Megan M. Burkhardt-Reed, Edina R. Bene, and D. Kimbrough Oller. Chapter 4 is in submission to the *Journal of Autism and Developmental Disorders* for publication. Its authors are Helen L. Long, Gordon Ramsay, Dale D. Bowman, Megan M. Burkhardt-Reed, and D. Kimbrough Oller.

Abstract

Long, Helen L. PhD. The University of Memphis. December 2020. Endogenous and social factors influencing infant vocalizations as fitness signals. Major Professor: D. Kimbrough Oller, PhD.

This dissertation evaluated the role of social and endogenous prelinguistic vocalizations as fitness signals in human development. It consists of three studies. The first investigated the reliability of listener judgments of the degree of infant vocal imitativeness in parent-infant vocal turn pairs as a measure of the saliency of potential vocal fitness signals. Participating listeners demonstrated moderate to high intra- and inter-rater agreement, suggesting vocal imitation has the potential to be used as a signal of fitness to caregivers in early development. The work also showed that vocal imitation in infancy is rare. The second study quantified the extent to which infants produce vocalizations socially (directed to a caregiver) vs endogenously (not directed to a caregiver) in laboratory settings where parents either attempted to engage them or talked with another adult. The infants produced three times as many vocalizations endogenously as socially in both circumstances. High rates of endogenously produced sounds may result from evolutionary pressures to signal wellness to caregivers through vocalization. Extensive independent vocal play may offer infants the opportunity to explore sensorimotor characteristics of the vocal system and provide the raw material that parents can use in face-to-face interactions. The third study examined social and endogenous motivations in the emergence of advanced vocal forms. Specifically, it compared canonical babbling ratios of infants at low and high risk for autism across high and low levels of both vocal turn taking and vocal play. Both groups showed a tendency to produce more canonical babbling during high turn taking and high vocal

play. The findings highlight a potentially robust internal social motivation for vocalization, even in the presence of likely social-cognitive differences such as risk for autism. High rates of endogenously produced canonical syllables in high-risk infants support the idea of robust evolutionary pressures for infants to signal fitness through vocalization. Furthermore, differences in vocal production across contexts can inform our understanding of the importance of both vocal interaction and independent infant exploration of vocalization. This dissertation offers perspective on the ways in which social and endogenous factors reveal natural selection pressures on fitness signaling in the human infant.

Table of Contents

List of Tables

List of Figures

1. Introduction

The stages of human development can inform our understanding of the selection pressures that differentiated us from our ape relatives (Griebel & Oller, 2008; Oller et al., 2016; Oller & Griebel, 2005, 2008). Using an evolutionary-developmental biology (evo-devo) framework, I follow the line of thinking that the stages of prelinguistic vocal abilities follow a natural logic of development that are foundational to advanced linguistic skills in humans. The ability to communicate using spoken language requires the ability to produce flexible vocalization that is not bound to any function or type of information and can be used with a variety of illocutionary functions (Austin, 1962) and varying emotional expressions (Jhang & Oller, 2017). These speech-like sounds, or "protophones" (Oller, 2000), can be used to regulate social interaction, share states of arousal, and explore vocalization itself from birth (Oller et al., 2013; Oller, Griebel, et al., 2019). Both human and bonobo infants produce fixed signals; however, ape infants produce fewer vocalizations with acoustically-similar features resembling those of human infant protophones without clear evidence of functional flexibility (Oller, Griebel, et al., 2019). These differences suggest an early distinction from our ape relatives in the capacity to develop language even by 2 months. My research is founded in the notion that hominin infants were more altricial than their ape relatives (Locke & Bogin, 2006; Robson et al., 2006), and thus under heightened selection pressure to signal wellness (Long et al., 2020; Oller, Griebel, et al., 2019).

I reasoned that infant protophones continue even today to be under selection pressure as fitness signals in human infancy. A reliable fitness signal used by infants would need to be salient and consistently perceived. Listeners must be able to judge infant vocalizations in terms of speech-like quality, level of distress, and the degree to which they conform to utterances produced by caregivers themselves (i.e., matching the quality of adult utterances). A key

selection force on vocal imitation is based on the fact that it can be easily interpreted by parents as a potential indicator of well-being. Thus, the first study in this dissertation in Chapter 2 will evaluate listener judgments on the level of "imitativeness" of infant vocalizations following parent models as a measure of the salience of the vocal signal. High inter- and intra-rater agreement on judgments of levels of imitativeness would suggest imitation is highly salient and has the potential to be used to signal wellness and general development (as it relates to vocal capabilities) to caregivers.

If infant vocalizations have the potential to signal fitness during development, it would be reasonable to assume that infants may experience greater pressure to vocalize more during face-to-face interaction, when the infant has the full attention of the parent. There is much research supporting the claim that parental interaction affects infant vocal production (Bourvis et al., 2018; Elmlinger et al., 2019; Franklin et al., 2013; Goldstein et al., 2003, 2009; Goldstein & Schwade, 2008; Gros-Louis et al., 2006), but there is also a growing body of evidence in support of intrinsic motivations to produce sounds for the infant's own purposes, i.e., endogenously (Moulin-Frier et al., 2014; Moulin-Frier & Oudeyer, 2013; Oller, Griebel, et al., 2019), in exploration of the sensorimotor characteristics of the vocal system. These two bodies of research would suggest that infants may produce more social protophones when socially engaged and more endogenous protophones during independent vocal play. The second study in Chapter 3 quantified the proportions of infant protophones perceived by listeners as either having a social or endogenous function to offer perspective on the relative roles of interactive and endogenous factors in infant vocal development throughout the first year of life.

Chapter 4 (Long et al., in submission) examined social and endogenous motivations involved in the emergence of canonical babbling across circumstances in infants at low and high

risk for autism. This final component of my dissertation aimed to address 1) how advanced forms of infant vocalizations can be used as fitness signals, and 2) the role of social and endogenous motivations for fitness signaling. The stable production of canonical syllables (i.e., adult-like consonant-vowel syllables with well-formed transitions) is a robust stage of development in the second half year of life, and parents are known to be reliable observers of their infants' onset of the canonical babbling stage (Oller et al., 2001). The onset of this stage can thus be considered a salient signal of developmental fitness and has the potential to illuminate social and endogenous motivations in infant vocal development and the foundations of language.

The study in Chapter 4 compared rates of canonical syllables across periods of high and low vocal turn taking and high and low independent vocal play in segments extracted from all-day recordings of infants in infants at low and high risk for autism. Autism spectrum disorder is a social communication disorder characterized by reduced social communication skills and by the presence of restricted and repetitive behaviors (American Psychiatric Association, 2013). The inclusion of autism risk groups was used to further elucidate the role of social motivation for fitness signaling. I proposed that positive selection pressure existed on the production of canonical syllables during social interaction for fitness signaling in typical development which may be absent or reduced in autism. Conversely, high-risk infants may present with more vocal repetition and self-stimulatory vocal behaviors during bouts of independent vocal play resulting in higher rates of canonical syllables produced when alone compared to the low-risk group. Lower rates of canonical syllables in the low-risk group may suggest these infants tend to explore the full range of sensorimotor aspects of the vocal system in support of vocal learning. Clinical group comparisons in infancy may also assist in identifying early predictors of impairments in children at risk for communication disorders such as autism. The findings

discussed throughout this dissertation emphasize the infant as an agent in vocal learning, offering perspective on the foundations for language.

2. Reliability of Listener Judgments of Infant Vocal Imitation (Long et al., 2019)

Abstract

There are many theories surrounding infant imitation; however, there is no research to our knowledge evaluating the reliability of listener perception of vocal imitation in prelinguistic infants. This paper evaluates intra- and inter-rater judgments on the degree of "imitativeness" in utterances of infants below 12 months of age. 18 listeners were presented audio segments selected from naturalistic recordings to represent in each case a parent vocal model followed by an infant utterance ranging from low to high degrees of imitativeness. The naturalistic data suggested vocal imitation occurred rarely across the first year, but strong intra- and inter-rater correlations were found for judgments of imitativeness. Our results suggest salience of the infant's vocal imitation despite its rare occurrence as well as active perception by listeners of the imitative signal. We discuss infant vocal imitation as a potential signal of well-being as perceived by caregivers.

Introduction

Imitation has been widely studied in infant and child development (Imafuku et al., 2019; Jones, 2007; Kugiumutzakis, 1999; Meltzoff, 1988a, 1988b; Meltzoff & Moore, 1977). Generally, the goal has been to seek insight about infant and children learning through imitation, with language learning being a special topic of interest (Bloom et al., 1974; Clark, 1977; Leonard et al., 1979; Moerk & Moerk, 1979; Rodgon & Kurdek, 1977). We have found no dispute in the child development literature regarding the importance of infant abilities to imitate as a foundation for language learning. But obvious instances of immediate imitation by infants of caregiver vocalizations do not occur very often (Papoušek & Papoušek, 1989; Pawlby, 1977; Užgiris et al., 1989). This raises the question of the possible importance of imitation by infants to parents in their understanding of the emergence of language in their children. To our knowledge

no prior research has addressed the possible importance of parental awareness of vocal imitation by their infants.

We reason that in spite of the low rate of vocal imitation, caregivers are aware of infant *abilities* to imitate because imitation may constitute an important signal of the infant's learning and well-being whenever it does occur. Thus, we are studying the sense in which vocal imitation may be a fitness signal to caregivers. Specifically, we seek to better understand infant vocal imitation as a signal occurring in naturalistic interactions by using a continuous rating scale to assess adult listeners' perceptions of the imitativeness of infant vocalizations. By examining imitation in this way, we assess the reliability of infants' use of imitation as a vocal signal of their developmental status.

Background

It is often claimed that babies learn language through imitation (Arbib et al., 2008; L. Bloom et al., 1974; Ghazanfar, 2013; Kugiumutzakis, 1999; Lewis, 1936; Mowrer, 1960; Schreibman, 2005). Others believe that infant imitation is present from birth as a way to map the actions of others who are "like me" onto a representation of their own actions to understand the psychological states of others and the self (Meltzoff, 2005, 2007) via active intermodal mapping (AIM) (Meltzoff & Moore, 1997, 2002) or via a mirror neuron system (Gallese & Goldman, 1998; Rizzolatti & Craighero, 2004; Simpson et al., 2015). These issues surrounding theories on the mechanisms and utility of infant imitation have been reviewed recently (Hurley & Chater, 2005a, 2005b; Jones, 2009; Keven & Akins, 2016; Oostenbroek et al., 2013; Ray & Heyes, 2011). In this study, we do not seek to redefine or rediscover the mechanisms involved in the utility of infant imitation; rather, we seek to assess the salience of the infant's imitation as a signal for caregivers from an evolutionary developmental perspective.

Experimental studies make up the majority of research testing infants' capability to produce imitation, with the focus largely on imitation of facial gestures (Heimann et al., 1989, 2017; Kuhl & Meltzoff, 1996; Meltzoff, 1988b; Meltzoff & Moore, 1977, 1983). However, we know of no empirical evidence on the capacity for listeners to make consistent judgments about the degree of imitativeness of individual acts. The only data we know of on subjective judgments of infant imitation have been dichotomous ratings in experimental studies for the purposes of assessing coder reliability (Barr et al., 1996; Carpenter et al., 1998; Collie & Hayne, 1999; Klein & Meltzoff, 1999; Meltzoff, 1988a, 1988b; Meltzoff & Moore, 1983, 1989; Sakkalou et al., 2013). This approach suggests that imitation is an all or nothing, binary skill. Our research will provide evidence of gradations in the extent of infant imitativeness and of the human listener ability to recognize such gradations.

Observational studies of infant vocal imitation have further provided an assessment of the frequency of imitation in parent-infant interactions (Masur, 2006). These, as well as experimental studies, require collecting subjective judgments on whether vocal acts are imitative (Užgiris, 2010). The occurrence of infant vocal imitation between ages 2 and 12 months in observational studies has been found to be low, occurring at <1 imitative event per minute (Papoušek & Papoušek, 1989; Pawlby, 1977; Užgiris et al., 1989). It is important to note that these studies have identified instances of imitation using different criteria: Užgiris & Pawlby (p. 111) reported judgments of imitation on the basis of the *totality* of utterances, and "not on the basis of specific aspects such as pitch" (Pawlby, 1977; Užgiris et al., 1989); in contrast, Papoušek & Papoušek evaluated imitative utterances by acoustic characteristics (i.e., pitch, duration, rhythm, and vowel or consonant resonance) and may have paid greater attention to the degrees in which utterances could be deemed imitative, thus potentially increasing the likelihood

that utterances would be treated as imitative (Papoušek & Papoušek, 1989). However, these judgments remained binary. While dichotomous judgments of infant imitation may provide useful evidence on infant capability and frequency of occurrence, we find it necessary to assess the salience of imitation as a signal using listener judgments of degree of imitativeness.

An Evolutionary Developmental Perspective on Infant Imitation

Within an evolutionary-developmental perspective (Bertossa, 2011; Oller et al., 2016), we propose that a key selection force on infant vocal imitation is based on the fact that it can be interpreted by parents as an indication of infant well-being, or fitness. Fitness is defined as the extent to which a biological trait is functional across a range of environments (Darwin, 1859; Latta, 2010). A reliable fitness signal used by infants would need to be salient and consistently perceived by listeners.

We follow the line of thinking that language emerges continuously with foundational capabilities building on each other (Oller, 2000; Oller et al., 2013). Specifically, early developmental skills and behaviors such as spontaneous vocalizations in the first month of life are seen within our perspective as foundational in building more complex skills such as canonical babbling and the infant's first words. The ability to imitate is also clearly foundational because learning to produce words requires being able to store and replicate phonological information. Thus, we seek to treat imitation as a feature of the emergence of language, recognizing that infant utterances can manifest varying degrees of imitation which can be interpreted by the caregiver as indicators of infant status in language learning.

Given that infant vocal imitation is infrequent, it would seem that parents must be acute in their identification of imitative utterances in order to make use of the information at all. In our longitudinal research we have noticed that parents in interviews with staff sometimes indicate

sounds their infant can imitate, but we have not yet quantified these tendencies. Attentiveness to rarely occurring imitation events could suggest that parental attention to imitation ability was selected for through hominin evolution as an indicator of infant growth of the language capacity. This likelihood suggests it is important to empirically evaluate how reliably imitativeness is transmitted by the baby to potential caregivers. The potential importance of such work is also supported by widespread suggestions that the ability of infants to imitate is associated with positive language and cognitive development (Masur & Eichorst, 2002; Ramer, 1976; Réger, 1986; Snow, 1989; Sundqvist et al., 2016).

In spite of the existence of numerous studies of vocal imitation and its importance in predicting language development in infancy, there has never been any prior study of infant vocal imitation to our knowledge that has attempted to establish a "gold standard" for judgment of infant vocal imitation. Nor has any research to our knowledge addressed what acoustic properties of matching between parent-modeled and infant-responsive utterances would influence degrees of perceived imitation. Yet it seems undeniable that human adults *can* make judgments about infant and child vocal imitation—the key empirical questions are 1) to what extent would listeners agree with each other if they did make judgments of imitativeness when presented with paired parent-infant vocalizations, and 2) to what extent would they be consistent in their own judgments if they made them repeatedly?

To provide empirical answers to these questions is the primary goal of this paper. We consider such work to be prerequisite to establishing standards of judgment about the nature of infant vocal imitation and a requirement for the development of ultimate gold standards for other research involving observational judgments of imitation. We take an evolutionary perspective wherein it is assumed that human caregivers and potential human caregivers *must* be able to

judge the vocalizations of human infants in terms of such issues as their speech-like quality, the degree to which they express distress, and the degree to which they conform to utterances produced by caregivers themselves (that is, the degree to which they are imitative). These abilities of caregivers, in accord with this evolutionary perspective, must be naturally selected because caregivers without such capabilities would be at a disadvantage in rearing successful children to compete for survival and reproduction. Thus, it seems that any normal[1] human adult must be able to judge infant vocal imitativeness to some degree. We started our empirical work for this paper with the assumption that such a capability would likely be present in any listener-participant with normal intellect.

How could we empirically evaluate such a capability? An obvious method is testing for inter- and intra-rater agreement on a substantial number of utterance pairs selected on an intuitive basis as showing a wide range of infant imitativeness. We reasoned that if any individual rater's judgments failed to show significant correlation with the ratings of a group of other persons, that individual would have been revealed as incapable of (or extremely poor in) judging imitation. The magnitude of observed correlations among raters would be reflective of the extent to which natural selection had yielded a strong signal of imitativeness in infant vocalizations as well as a strong capability in listeners to recognize that signal.

The evolutionary perspective also suggests that although we do not know what magnitude of agreement to expect among and within listeners, we can expect statistically significant agreement. As argued above, a human who is unable to recognize vocal imitation would be at a

[1] From an evolutionary biology perspective, "normal" refers to the statistical distribution of biological traits and cultural views about these traits on what bodies "should" be like, also known as "biological normalcy" (Wiley & Allen, 2017; Wiley & Cullen, 2020). In general a "normal" individual has the potential to survive and reproduce.

disadvantage in recognizing all aspects of speech signals, indeed would not likely be able to understand speech, nor to judge the content of vocalizations of babies. Such a person would be at a disadvantage in trying to make sense of the vocal communications of their own progeny, and the progeny would presumably experience negative selection pressure due to concomitant insensitive parenting. We reason thus, that after many generations of selection, all persons without any significant capability to judge imitation would have been weeded out.

While it is expected that all raters will be significantly able to judge imitativeness (i.e., would show significant agreement with other raters), the evolutionary perspective also predicts that there must be variation both within and among raters—all traits that are subject to natural selection must show variation (Darwin, 1859; Locke, 2009; West-Eberhard, 2003). Evolutionary theory therefore suggests we should attend to variation both within and across observers.

A key point about such research is that there is, at present, no basis for asserting a "gold standard" for judgment of imitative and non-imitative events. Although we assume all normal humans should be able to significantly judge imitation, how would we know that one person is better at it than another? Even following significant experience in working with and making judgments on imitation, there would be no empirical way to assess that a person is particularly good at judging imitation in the absence of a measure of that person's agreement with the standard of humanity in general on judgements of imitation. Thus, we presume that research determining agreement within and across a panel of normal human listeners is a prerequisite to the establishment of an empirical gold standard for judgments of imitation and for providing empirical perspective on the role of imitation as a salient fitness signal to caregivers.

Methods

Data Collection

Approval for the longitudinal research that produced data for this study was obtained from the University of Memphis Institutional Review Board for the Protection of Human Subjects. Data were acquired from archives of the longitudinal investigations on typically developing infants in and around Memphis, Tennessee, and all parents spoke English in the selected laboratory recordings. Recruitment for this archival data was conducted in child-birth education classes and by word of mouth. Parents or prospective parents of newborn infants were presented with a detailed consent form after having been interviewed as possible participants in the longitudinal recordings. One infant was exposed to Ukrainian and English at home, but all other infants were exposed to only English at home. Criteria for inclusion of infant participants included a lack of impairments of hearing, vision, language, or other developmental disorders.[2]

We drew from archived audiovisual recordings of six parent-infant dyads (3 male, 3 female infants) representing naturalistic interactions in a laboratory setting. During recordings, the parent-infant pairs occupied a studio designed as a child playroom with toys and books. Laboratory staff operated four or eight pan-tilt video cameras located in the corners of a recording room from an adjacent control room—there were three such recording laboratories at varying stages of the research. In all the laboratories, two channels of video were selected at each moment in time with the goal of recording 1) a full view of the interaction and 2) a close view of

[2] Because parents were recruited during pregnancy, inclusion criteria for participation was initially determined as a normal pregnancy up to the point of recruitment without any detected complications. Typical development of the infant for later analysis was confirmed throughout participation in the longitudinal study via parent report during laboratory visits using information such as passed hearing screenings and mastery of developmental milestones at approximately expected ages.

the infant's face. Both the parent and the infant wore high fidelity wireless microphones, with the infant microphone <10 cm from the infant's mouth. Detailed descriptive information regarding laboratory equipment used can be found in previous studies completed from this laboratory (Buder et al., 2008; Warlaumont, Oller, Buder, et al., 2010).

Two laboratory recording sessions were selected from all 6 infants at approximately 3, 6, and 10 months, for a total of 36 recordings used to select utterances. The average length of sessions used for this study was 19 minutes (range: 12-22 minutes). These sessions were selected from longer recordings which often lasted around 60 minutes, during which parents were asked to interact with their infant or with a laboratory staff member. Demographics and recording age for each infant at each session are tabulated in Appendix A.

Identifying Functions of Infant Vocalizations

All infant vocalizations across the recordings were initially labeled in terms of illocutionary force, defined as potentially communicative functions of the utterances (Austin, 1962; Oller et al., 2016; Searle, 1969). We sought all possible instances of imitation, which was one of the illocutionary forces coded, in both interactive or non-interactive contexts throughout the recordings we examined. The coding was done within the Action Analysis Coding and Training software (AACT) (Delgado et al., 2010), used and discussed in previous research from this laboratory (Jhang & Oller, 2017; Warlaumont, Oller, Buder, et al., 2010; Yoo et al., 2018). Pre-linguistic infants express varying emotional content (i.e., positive, neutral, and negative) in early vocalizations beginning at birth (Jhang & Oller, 2017; Oller et al., 2013). Infants have been shown to have the capacity to produce a single vocal type with multiple illocutionary forces on different occasions, suggesting they possess the foundations necessary for the variable illocutions seen for words and sentences in mature language. Following this thinking, pre-

linguistic infant vocalizations can be used during dyadic interaction with varying communicative intentions, or vocalizations can be internally driven and produced for the infant's own purposes. Viewed in this context, vocal imitation is a kind of illocution, a function performed when an infant produces a sound that reveals matching to a heard sound.

Imitative vocalizations were coded as exhibiting any degree of infant imitation as observed by the adult listener (author 1) who selected the stimuli, taking into account auditory and acoustic characteristics such as matching pitch contour, number of syllables, and/or syllable types in both dyadic and non-dyadic contexts. A non-dyadic circumstance could be, for example, if an infant imitated a caregiver who was not talking to the infant but offering examples of infant utterances to a laboratory interviewer. A total of 6,474 utterances were labeled for illocutionary force in the 36 recordings used in this study.

Extraction of Stimulus Pairs

Our goal in stimulus selection was to acquire a set of infant vocalizations that represented the broad continuum from high imitativeness to no imitativeness from the 6,474 utterances. We do not assume that there exists a gold standard for categorizing infant utterances into three groups of high, low and no imitativeness, but we aimed to select utterances roughly equally in these three intuitively determined groups in order to ensure that we would have stimuli across the entire continuum. The groups were used as a heuristic for the selection process and were not theoretically important, so we did not endeavor to make the selections precisely equal in the three groups.

Only 299 infant utterances were identified as showing *any* degree of imitativeness, less than 5% of all the utterances in the recordings.[3] From these, 108 utterances along with the preceding parent utterances, were selected to be extracted and used as stimulus items for listener judgments. 60 of these were designated intuitively as showing "low" imitativeness and 48 as showing "high" imitativeness. The remainder of the 299 imitative utterances were eliminated because they 1) had a low signal-to-noise ratio, 2) poor recording quality, 3) high parent-infant voice overlap, 4) repeated imitations (without repeated preceding adult models), or 5) speech occurring between the model and the imitation. 58 additional pairs were identified from the original 6,474 in the recordings as clearly not imitative and were extracted for the purposes of including non-imitative infant utterances in the stimulus set, as long as these utterance pairs were not disqualified by any of the 5 elimination criteria above. This procedure ensured a wide range of possible judgments on degree of imitation. A total of 166 stimulus pairs were therefore used for listener judgments. Figure 1 in Appendix B provides a visualization of the flow for the selection of stimulus pairs and additional commentary. Also, in Table 4 of Appendix D we provide 10 example stimuli wav files used in this experiment.

Listeners and Rating Scale

Eighteen listeners were asked to rate the degree of imitation for each of the 166 pairs. The participants included 15 graduate assistants (MA, AuD, and PhD graduate students in the School of Communication Sciences and Disorders) and 3 staff members of the Origin of

[3] A second observer, blind to the purposes of the study, coded 11 recordings (30%), which had been selected at random from among the 36. A correlation of 0.88 was found across the 11 recordings between the primary and secondary observer on number of imitative utterances designated. The outcome for both coders on the selected recordings conformed to the widely reported tendency for vocal imitation to be found to occur rarely in infancy (see citations above), and in fact the second observer coded less than 2/3 as many items as imitative (16) as the primary coder (25) across the 11 recordings.

Language (IVOC) Laboratory, all of whom were female. The listeners had no previous experience rating infant utterances on a continuous scale or making judgments of degree of imitativeness; however, all listeners had experience listening to infant sounds and identifying vocalization types (e.g., squeals, growls, vowel-like sounds, etc.) and canonical syllables.[4] The first author, who selected the stimulus pairs, also participated as a listener (hereafter, "Rater 1").

Rating Scale

A continuous rating scale (range 0-100) was presented to listeners in the AACT software environment (Delgado et al., 2010) for making judgments on the degree of imitativeness of infant utterances as compared to adult models. See Figure 2 in Appendix C, which provides a screen shot of the scale tool. The listeners, prior to hearing any of the stimuli, were shown a screen shot of the rating tool and it was explained to them that when using the tool they would merely click with a mouse pointer on any location within the scale each time they would hear a stimulus, and AACT would assign a number from 0-100 indicating the degree of imitativeness specified. Listeners were encouraged to use the entire scale.[5] The scaling tool was very easy to use, and none of the raters expressed any difficulty in managing the rating task.

[4] Four raters (Raters 1, 2, 10, and 11) had previous coding experience identifying social and non-social functions of infant utterances, including a category labeled *Imitation.* However, training for this category included only the brief presentation of a list of auditory-perceptual criteria to consider for imitation. The raters were instructed to make their judgments based on intuition. Rater 1, who selected the stimuli, and worked closely with the last author, was the only member of the group that could be thought to have engaged in a sort of training on imitation. Raters were between 21-40 years of age, two were parents, and all had at least a bachelor's degree.

[5] One listener reported selecting a "Show Rating" option that was available on the rating scale, which resulted in a display of the digital value (0-100) associated with the position on the visual scale for each placement of the cursor. The remaining listeners did not see the numerical values.

Instructions for Listeners

Listeners were presented minimal instructions on how to make judgments of infant imitativeness. Specifically, they were told to broadly consider auditory-acoustic characteristics such as duration, pitch, syllabicity, and articulation of the parent and infant utterances when making their judgments. Our goal was to encourage listeners to use their natural intuitions about infant vocalization and thus hopefully for them to simulate mothers' judgments of imitativeness. Because these pairs were selected from infants younger than 12 months of age, listeners were encouraged to rate the degree of imitation regardless of whether the infant utterance was exactly like that of the caregiver (e.g., a word imitation).

Calibration Stimulus Pairs

In order to ensure listeners understood the task, 12 calibration pairs were selected from the 166 stimuli by the first author and presented prior to the judgment task as examples of very high (6 pairs) or very low (6 pairs) degrees of imitativeness within the sample. The calibration pairs were not rated by the listeners during this presentation. These pairs were also included and randomized in order within the stimulus set in the full listening judgment task. Mean ratings for the calibration items that were made by the listeners during the full judgment task, along with ratings for the other stimulus pairs, can be seen in the discussion on rating scale usage in Appendix C.

Listening Judgment Task (Rating Trials)

After listeners were presented instructions and the calibration stimulus pairs, the formal rating task began, with five randomized trial blocks of the 166 pairs presented to each listener. In other words, all 166 pairs (including the 12 calibration pairs) were presented five times to each listener for a total of 830 rating trials. The set of 166 stimulus pairs was randomized within each

trial block. The beginning and ends of each trial block were inspected to ensure no single pair was presented twice within 10 consecutive stimulus pairs. The task took approximately one and a half hours for each listener to complete.

Results

To ensure that the scale was being utilized appropriately, we first examined the range of ratings used by all the listeners. Zeroes occurred commonly in the ratings, and the highest minimum rating by any individual was 2 (mean minimum rating across all listeners: 0.3); ratings of 100 were also fairly common, and the lowest maximum rating by any individual listener was 97 (mean maximum: 99.4). Almost 2/3 of the ratings occurred in the middle of the scale from 20-75. All listeners were thus confirmed to have utilized essentially the entire scale for their judgments. See Appendix C for graphic analyses of rating scale usage and mean rater bias.

Inter-Rater Correlations

To compute mean inter-rater correlations (MICs), we first calculated the mean rating across the 5 trials on each stimulus pair for each listener. We will refer to these as the individual rater means (IRMs). We paired the IRM for each stimulus and for each rater with the IRMs of all the other raters and computed the 17 correlations for the pairings (n =166). An MIC was calculated for each rater across these 17 pairings, and each of these MICs is represented in Figure 1 as a red diamond. The mean of the MICs, 0.71 (range: 0.66 to 0.76, n = 166 for each), was highly significant, $p < .00001$ (SD across the 18 MICs = 0.03, 95% CI [0.72, 0.69]). Even the lowest of these inter-rater correlations was highly significant ($p < .00001$, n = 166). These mean inter-rater correlations suggest moderate to strong positive relationships across raters for judgments on each stimulus pair, as expected based on the assumption that all normal human listeners should have an evolved capacity for recognizing vocal imitation. Although the

agreement among listeners was highly significant, it is also true that the listeners showed notable and often significant differences from each other in the degree to which they agreed with the other listeners, as indicated by the error bars (95% CIs) of the MICs in Figure 1.

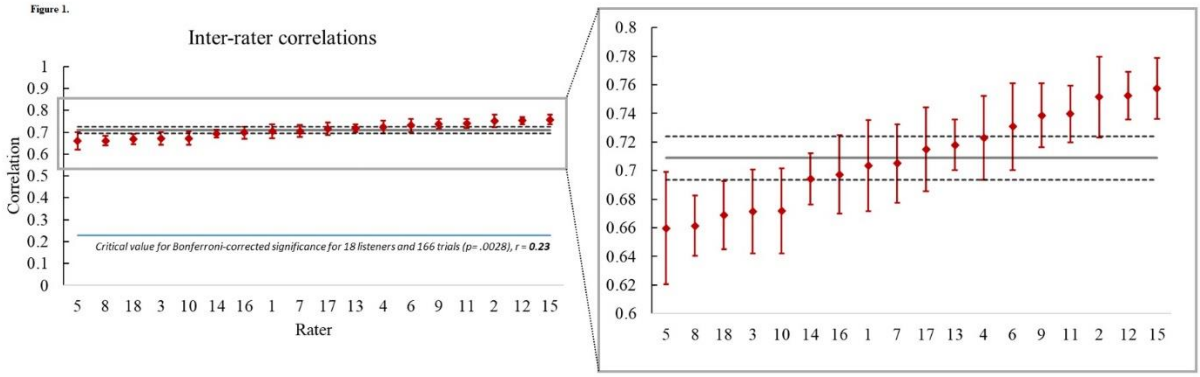

Figure 1. Mean inter-rater correlations
Mean inter-rater correlations (MICs) ordered from lowest (r = 0.66) to highest (r = 0.76) for each of the 18 raters represented by red diamonds. On the left, the entire correlational scale is represented, and on the right a blowup is offered for the region where the scores occurred. On both the left and right, the solid horizontal gray line represents the mean (0.71) across the 18 MICs and the dotted gray lines correspond to the 95% confidence interval (0.69–0.72) for that mean. On the left we also show, with a horizontal blue line, the critical value for statistical significance of the correlations; the huge gap between the critical value correlation and the actual correlations makes clear that the ratings of all the listeners were correlated at a highly significant level ($p < 0.00001$) with those of the other raters. At the same time the blowup on the right makes it possible to easily examine differences among the 18 listeners in their levels of agreement with the other listeners by evaluating the means and 95% CIs (the error bars) for any pair of listeners. For example, Rater 5 agreed significantly less with the others than Rater 15, since their CIs do not overlap at all. To compare any two raters' levels of agreement with that of any of the other raters, observe the error bars of one with respect to the mean of the other; if the CIs for the first rater do not overlap with the mean for the other rater, the two are significantly different at $p < 0.05$.

Intra-Rater Correlations

Intra-rater correlations, which reflect listener consistency of rating across trial blocks, were calculated from the mean within-rater correlations of the 166 IRMs across each of the 5 trials on each stimulus; all 10 possible pairings of the five trials for each listener were correlated. The average intra-rater correlation was $r = 0.73$ ($p < .00001$ for all ratings, SD = 0.08, 95% CI = .69, .77) ranging from 0.54 to 0.84. These results suggest moderate to strong positive relationships between individual rater judgments of each stimulus pair across trial blocks as

shown in Figure 2, with the mean within-rater correlation for each listener represented as a black diamond.

As with inter-rater agreement, the very significant intra-rater agreement was also accompanied by differences among the raters in the degree to which they showed consistency in rating across the five trials. These differences are again reflected in means and CIs for the individual listeners in the figure.

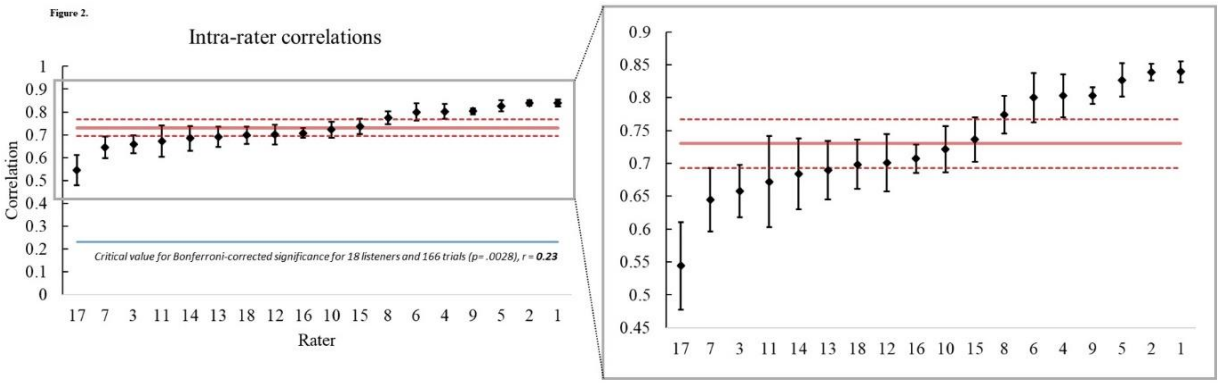

Figure 2. Mean intra-rater correlations
Mean intra-rater correlations organized from lowest (r = 0.54) to highest (r = 0.84) for all 18 listeners, with an average correlation of 0.73 indicated by the solid red line with a 95% CI of 0.69 −0.77 represented by dotted red lines. Intra-rater correlations were calculated for each of the 18 listeners by averaging the correlations of their ratings across all possible pairings of the five trial blocks for each of the listeners. The left-right distinction is as in Figure 2. Again, on the left there is a huge gap between the critical value correlation (blue line) and the actual correlations across the 18 listeners, making clear that all the intra-rater correlations were highly significant ($p < 0.00001$). Again, the blowup on the right makes it possible to easily examine differences among the 18 listeners in their levels of agreement with their own ratings across the 5 trial blocks, i.e., their rating consistency.

Intra-Rater Bias: Change in Ratings Over Trials

We also evaluated the statistical significance of the within-rater differences in rating levels between individual trials. Unlike the intra-rater correlation, this analysis compares each listeners' rating levels, comparing those levels for each trial block with all the other trial blocks (again in all 10 possible pairings), providing information on how raters changed their rating biases over time (e.g., higher or lower average ratings trial to trial).

Column 1 in Table 1 indicates the rater, and the subsequent columns indicate the *p*-values for the 10 possible pairings across 5 trials for each stimulus pair. A Kolmogorov-Smirnov test was conducted for each pairing. This is a non-parametric test of distributions, which can detect differences in mean ratings across trials or differences in distribution shape across trials. The null hypothesis for this test was that the ratings from the two trials came from the same distribution. In other words, a *p*-value >.05 indicates the two paired trials were not significantly different from each other. For example, the ratings from the first two trials for Rater 1 were not significantly different from each other ($p = .349$). The first and third, on the other hand, were in fact significantly different ($p = .006$). The third, fourth, and fifth ratings were not statistically different, $p = .779, .689, .507$, respectively. Three of the 10 pairings for Rater 1 showed statistically significant differences of ratings.

Table 1. Agreement across judgments within raters

There were 119 out of a total of 180 comparisons with non-significant differences (p-value < .05). In other words, 61% of the comparisons show raters were overall consistent in their judgments, whereas the remaining 39% suggest raters changed their decision patterns across trials. A 2x2 chi-square test of independence determined that this pattern of listener changes across trials occurred at a rate much greater rate than chance, $\chi 2(9) = 143.32$, $p < .001$.

	p-value of test of agreement across judgments within rater									
Rater	**1-2**	**1-3**	**1-4**	**1-5**	**2-3**	**2-4**	**2-5**	**3-4**	**3-5**	**4-5**
1	0.349	0.006	0.013	0.009	0.108	0.108	0.108	0.779	0.689	0.507
2	0.779	0.283	0.424	0.083	0.859	0.991	0.180	0.968	0.018	0.14
3	0.108	<.001	0.002	0.034	0.062	0.349	0.689	0.083	0.002	0.507
4	0.924	0.424	0.227	0.018	0.779	0.507	0.062	0.998	0.507	0.507
5	0.283	0.140	0.001	0.083	0.083	<.001	<.001	0.507	0.108	0.689
6	0.227	0.140	0.062	0.108	0.991	0.859	0.689	0.689	0.968	0.349
7	0.013	0.006	0.507	0.596	0.779	0.349	0.083	0.108	0.013	0.689
8	0.025	0.227	0.083	0.083	0.968	0.046	0.034	0.349	0.283	0.779
9	0.507	0.424	0.004	0.004	0.006	<.001	<.001	0.424	0.424	0.924
10	0.596	0.068	0.689	0.013	0.09	0.968	0.018	0.284	0.848	0.227
11	0.859	0.006	<.001	<.001	0.034	<.001	<.001	0.006	0.001	0.859
12	0.859	0.227	0.924	0.083	0.596	0.998	0.227	0.349	0.779	0.227
13	0.018	0.006	0.025	0.227	0.001	0.227	0.025	0.046	0.001	0.034
14	0.779	0.283	<.001	<.001	0.507	0.009	0.001	0.083	0.034	0.424
15	0.001	0.108	0.002	0.083	0.507	0.14	0.283	0.424	0.596	0.596
16	0.596	0.424	0.507	0.859	0.859	0.968	0.859	0.779	0.779	0.968
17	0.924	0.689	0.003	0.034	0.424	0.002	0.034	0.006	0.002	0.046
18	0.349	0.001	<.001	<.001	0.108	<.001	0.034	0.034	0.507	0.227

Discussion

The primary finding based on these data is that listeners were consistent both within their own repeated judgments and with other listeners on ratings of the degree of imitativeness in infant vocalizations from three to twelve months. Judgments of utterances inclusive of a wide range of imitativeness and lack of it evidenced significant moderate to strong relationships within and across raters, and these differences were highly significant statistically. The raters actually judged very few utterances as highly imitative—despite 48 out of the 166 pairs having been initially selected as being "highly imitative"—with only 5% of the mean ratings for the 166

pairs of parent and infant utterance exceeding 80 on the 100-point scale. Yet, the significant moderate to strong correlations indicate salience to the listeners of the imitative signal, even though it appears to have been weak. These results lead us to speculate that vocal imitation in the first year of life is a trait that may have undergone positive selection pressure as a fitness signal to indicate communicative well-being of the human infant.

The importance of the reliability of infant imitation as a signal seems augmented by the fact that imitation was observed rarely across the 36 recordings from which the stimulus materials were drawn. We found only 299 instances of utterance pairs where any degree of imitativeness was perceived by the stimulus selector out of 6,474 total infant utterances. These results are consistent with previous findings reporting that infant vocal imitation in naturalistic interactions does not occur frequently (Papoušek & Papoušek, 1989; Pawlby, 1977; Užgiris et al., 1989).

All in all, the results support an interpretation of the perception of infant vocal imitation that emphasizes salience of the imitation signal, as indicated by highly significant correlations among and within raters on judgments of utterances with regard to imitativeness. This salience suggests vocal imitation, though infrequent in occurrence, may serve as a fitness signal with regard to infant communicative abilities.

At the same time, the perception of the imitative signal shows variation in salience across different listeners as well as changes across time in judgments made within individual listeners. Trait variation among conspecifics is a primary postulate of Darwin's theory of evolution by natural selection (Darwin, 1859; Latta, 2010). The interpretation invokes the two evolutionarily necessary sides of imitativeness as an evolving trait: on the one hand it must show a measure of stability—reflected in fairly consistent perceptions of it—while on the other hand there must be

variability in its perception, because without that, there would be no potential for natural selection of imitation as a fitness-signaling trait.[6]

Limitations and Future Directions

A limitation of this work relates to the small number of listeners as well as the selection of them, all being female, living in the USA. Also, all of them were associated with the IVOC laboratory with some experience identifying categories of infant sounds, but importantly with *no* experience rating degrees of imitativeness prior to participating in the study.[7] Thus, our results cannot be generalized to all possible human listeners.

We have little reason to think the laboratory training that had been involved, namely training in infant vocalizations coding, had notable influence in our study. Four of the 18 listeners had engaged in some coding that had required them to label infant utterances for illocutionary force (Austin, 1962; Oller et al., 2016) where one of the possible categories was "imitation". But again, these four raters showed correlations very much like those of the other listeners and showed correlations with each other that were typical of the group. Even the first author, who was one of those four, and who had selected the stimuli, showed a typical agreement level with the others. Another important potential expansion of this work would be to compare male and female listeners. There have been other cases where gender differences have been

[6] At the level of individual listeners, variation among raters may also be attributed to differences in each rater's level of attention to pairs across all 830 trials, individual auditory perceptual abilities (Arazi et al., 2017), or an individual's use of the rating scale (i.e., tendency to use the full range of the 0-100 scale or to only interact with certain areas of the scale such as the low or high extreme ends). Further evaluation of individual variability is necessary to better understand listener-related differences on perceptual judgment tasks.

[7] An additional limitation to this study is that there were not inclusion or exclusion criteria associated with neurocognitive functioning. It is possible that screening cognitive abilities may reveal differences among the listeners that could explain for some of the variation noted among listeners. Further evaluation assessing the relation between cognitive functioning and perceptions of acoustic-perceptual characteristics associated with vocal imitation are warranted for a deeper understanding of this topic.

found in perceptions of child development (Hastings et al., 2005; Kerig et al., 1993; Siegal, 1987), and consequently we cannot be sure that the patterns found here would apply equally to fathers or other male caregivers.

Future studies will hopefully assess listener differences by comparing experienced infant caregivers (individuals who have presumably made many tacit or explicit judgments about the imitativeness of infant vocalizations) with individuals having had little or no such experience. The two parents among the 18 listeners showed average rating agreement with the other listeners that was very near the mean for all the listeners, but because there were only two, we think further inquiry into a possible role for parenting experience is warranted. The experience of growing up in different cultures could also play a role, and we deem it important to evaluate judgments made by persons from different language and cultural backgrounds and presumably conditions of SES. Though we know of no research on vocal imitation rates being influenced by SES, there is a substantial literature on other kinds of SES effects in child development (Conger & Donnellan, 2007; Hoff-Ginsberg, 1998; Hoff, 2003).

Future directions for this line of research might also assess individual differences in rates of vocal imitation by infants. The current sample is too small (only six infants) to yield a persuasive picture on the matter, although the range of imitated utterances across the six infants was notable, from ≈ 22 per hour to ≈ 1 per hour (mean ≈ 10 per hour) in this sample of 6 recordings from each infant (see Table 2 in Appendix A). Similarly, the sample was too small to make much of gender differences, but the three girls had much higher rates (mean ≈ 16 per hour) than the three boys (mean ≈ 3 per hour). Although we know of no research on vocal imitation rates in naturalistic samples for boys and girls, there is of course a considerable literature base on gender differences in other realms of language development (Gleason & Ely, 2002; Huttenlocher

et al., 1991). Furthermore, the girls tended to have mothers with higher educational levels—a common indicator of SES—than the boys. Another issue is that all the girls were first-borns whereas only one of the boys was (and he had the highest imitation rate among the boys ≈ 8 per hour). Again, we know of no research on imitation rates being affected by birth order, but there is a substantial literature on birth order effects other realms of child development (Breland, 1974; Hoff-Ginsberg, 1998; Zajonc & Markus, 1975).

Whatever the individual caregiving experience, gender, cultural or SES effects are determined in the future to be across a broad range of infants, it would also be useful to assess parents' perceptions of their *own* infant's imitation skills. Parents in our laboratory have sometimes asserted that their infants imitate frequently, suggesting that imitation is a salient indicator of vocal development for those individuals. It would be useful to determine whether parental perceptions correspond to the actual infant rates of imitation or whether either the rates or the parental perceptions of them are predictive of later vocal development.

3. Social and Endogenous Infant Vocalizations (Long et al., 2020)

Abstract

Research on infant vocal development has provided notable insights into vocal interaction with caregivers, elucidating growth in foundations for language through parental elicitation and reaction to vocalizations. A role for infant vocalizations produced endogenously, potentially providing raw material for interaction and a basis for growth in the vocal capacity itself, has received less attention. We report that in laboratory recordings of infants and their parents, the bulk of infant speech-like vocalizations, or "protophones", were directed toward no one and instead appeared to be generated endogenously, mostly in exploration of vocal abilities. The tendency to predominantly produce protophones without directing them to others occurred both during periods when parents were instructed to interact with their infants *and* during periods when parents were occupied with an interviewer, with the infants in the room. The results emphasize the infant as an agent in vocal learning, even when not interacting socially and suggest an enhanced perspective on foundations for vocal language.

Introduction

The relative frequencies of human infant vocalizations that can be categorized as social vs. endogenous have not been a major focus of research. We seek to quantify the extent to which infants vocalize socially and endogenously in naturalistic settings. The effort has led to a shift in our perspective, where the contribution of endogenous vocalization and exploratory vocal play has assumed increasing importance in our speculations about the emergence of the speech capacity both in development and evolution.

The new perspective is informed by evolutionary developmental biology, evo-devo (Bertossa, 2011; Carroll, 2005; Müller & Newman, 2003; Newman, 2012), a paradigm of thought that emphasizes natural selection as targeting developmental processes, allowing the

evolution of foundational structures and capabilities upon which subsequent developments can self-organize and be further exploited in subsequent development and evolution. This approach does not diminish the importance of social interaction in the origin of the speech capacity, but instead is hoped to help account for foundational requirements of functionally flexible vocal interaction. In essence, the line of reasoning emphasizes the origin of flexible vocalization, without which significant growth in flexible vocal interaction and, through further development, vocal language may have been impossible.

Social Interaction and Vocal Development

The effect of social interaction on infant vocal development has long been a topic of interest in child psychology and the emergence of language (Bloom et al., 1987; Bloom & Esposito, 1975; Goldstein et al., 2009; Goldstein & Schwade, 2008; Gratier et al., 2015; Gros-Louis et al., 2014; Hsu & Fogel, 2001; Iyer et al., 2016; Lee et al., 2018). The study of infant intrinsic motivation for social engagement has highlighted an apparently innate drive to engage in face-to-face dyadic interaction with caregivers from birth (Trevarthen, 1979, 1998) and has been interpreted as contributing to the development of temporal sensitivity, vocal coordination, and social contingency (Crown et al., 2002; Roberta Michnick Golinkoff et al., 2015; Jaffe et al., 2001; Ramírez-Esparza et al., 2014; Roseberry et al., 2014). The long tradition of research in infant attachment and bonding (Ainsworth, 1969; Bowlby, 1969; Pipp & Harmon, 1987; Schore, 2001) has included a distinct emphasis on the parent-infant dyad as the fundamental unit of human social and emotional development. Even in the first 3 months of life parent-infant vocal interaction has been described in detail (Dominguez et al., 2016; Gratier & Devouche, 2011; Yoo et al., 2018). Experimental studies in the still-face paradigm (Tronick et al., 1978) have shown that by 5-6 months of age, infants increase their rate of speech-like

vocalizations when the parent disengages from an ongoing vocal interaction (Franklin et al., 2013; Goldstein et al., 2003), suggesting infants by that age seek to repair broken interactions with increased vocalization. A social feedback loop has been posited to exist in infant and child vocalization, and that loop has been thought to promote contingent infant vocalizations with respect to caregiver vocalizations (Abney et al., 2017; Gros-Louis et al., 2014; Hsu & Fogel, 2003; Warlaumont et al., 2014). Winnicott (Winnicott, 1960) went so far as to say that "there is no such thing as an infant," highlighting the idea that without a mother, an infant cannot exist. But this idea has been taken too far, we think, if it is interpreted to imply that research on human infancy should emphasize the dyad to the near exclusion of interest in the independent infant as an agent in its development.

There can be no doubt that social interaction plays a critical role in infant vocal learning and language acquisition; social learning allows us for example to acquire language-specific syllables, phonemic elements, and the largely arbitrary pairings of words with meanings in languages. But even deaf infants produce the same kinds of prelinguistic speech-like sounds, or "protophones" (Oller, 2001), as hearing infants in the first year of life (Oller & Eilers, 1988). Thus the importance of *hearing* speech sounds from the social environment does not appear to drive the initial development of protophones. In this paper, we seek to highlight the quantity of infant endogenous, non-cry vocal activity to further illuminate the role protophones play in supplying a basis for social learning.

Several studies have shown that dyadic vocal interaction increases the rate of protophone production (volubility), and the proportion of advanced vocal forms including canonical babbling appears to be particularly high during dyadic vocal interaction (Goldstein & Schwade, 2008; Gratier & Devouche, 2011; Gros-Louis et al., 2014; Hsu & Fogel, 2001; Lee et al., 2018). Yet

surprisingly, the proportion of infant protophones that are social in nature has, to our knowledge, never been previously quantified, so the extent to which infant protophone production may be primarily social rather than endogenous is unknown.

Intrinsic Motivation to Support Vocal Development

Intrinsic infant motivation for action and exploration has long been recognized. For example, Piaget's sensorimotor stage in the first two years of life is portrayed as a period wherein infants' self-generated gestures are produced without social intent, but rather for the pure enjoyment of experiencing sensorimotor activity (Piaget, 1952a, 1952b). In anecdotal reports (Caligiore et al., 2008; Grossberg & Vladusich, 2010; Pedersen et al., 1979; Sheya & Smith, 2013; Vauclair & Bard, 1983), the interpretation of this stage focused on the circular reactions of manual gestures, but Piaget did not emphasize circular reactions in the vocal domain (Stark, 1981).

The low level of focus on the infant as an independent agent of vocalization in prior research on development (see Appendix E) might be in part an unintended consequence of the radical behaviorist tradition that for many decades treated behaviors as *responses* rather than *actions* (Skinner, 1957; Watson, 1913). Panksepp and his colleagues have argued that we have not overcome the legacy of that radical behaviorism, and that even modern cognitive psychology continues to underplay the endogenous, emotion-driven actions of both humans and non-humans (Davis & Panksepp, 2018; Panksepp, 1982, 2011; Panksepp & Biven, 2012).

Breaking with the dominant tradition of infant development research, a role for intrinsic motivation as a primary mechanism to support vocal development has recently received increased attention (Moulin-Frier et al., 2014; Moulin-Frier & Oudeyer, 2013; Oller, Griebel, et al., 2019). In the Supplementary Material to a published article based on recordings made in our

own laboratory (Oller et al., 2013), it was reported that infants across the first year of life produced the majority of their protophones when gaze was not directed toward another person. In a small-scale study from another laboratory with just 16 minutes of recording per infant at 6-8 months, infants produced more vocalizations when playing alone with toys than when engaged socially (Harold & Barlow, 2013). Another recent observational study found no significant difference in protophone volubility between a recording circumstance where parents talked to infants compared to circumstances where parents were in the same room and silent or not present in the room at all, suggesting that infants had an "independent inclination to vocalize spontaneously" in the absence of social interaction (p. 481) (Iyer et al., 2016). Importantly, the rate of protophone production has been reported to be very high, >4 protophones per minute during all-day audio recordings, across the entire first year, and even when infants were judged to be alone in a room, the rate was >3 per minute (Oller, et al., 2019).

These findings suggest vocalizations are commonly produced endogenously. In other words, infants in these prior studies appear to have been intrinsically motivated to explore or practice sounds, in essence to play with sensorimotor aspects of sound production, although the evidence has been somewhat indirect. We propose that this vocal exploration may have a deeply significant role in vocal development, alongside the importance of caregiver-infant interaction and ambient language exposure. In spite of the possible importance of endogenous, exploratory vocalizations in language development, to our knowledge there is no published evidence specifically targeting the communicative function of infant protophones or the lack of it. Only with such work will it be possible to reliably quantify proportions of endogenous infant protophones and socially-directed ones. (see Appendix F, for information suggesting that both

parents and non-parents tend to view infant vocalizations as being predominantly social rather than endogenous or exploratory).

We deem it important that such quantification be established in contexts with and without parent engagement across the first year of life. Prior studies suggest the proportions of endogenously-produced sounds may be high, but appropriate research requires direct comparison in different circumstances of potential interaction, especially when caregivers are attempting to interact with infants and when not. Providing such quantification may highlight the importance of endogenously generated vocalization and self-organization in prelinguistic vocal development (Moulin-Frier et al., 2014; Moulin-Frier & Oudeyer, 2013) and may help establish perspective about relative roles of endogenous and interactive factors in vocal development.

Specific Aims and Hypothesis

Our primary goal is to determine the extent to which infants produce social and endogenous vocalizations at three ages and in two laboratory circumstances: An *Engaged* circumstance, where the parent attempts to interact with the infant, and an *Independent* circumstance, where the infant is present in a room, but the parent is interacting with another adult. This quantification is hoped to provide a standard against which we may be able to recognize the relative importance of infant protophones both as social and as endogenous. We hypothesize that infants will produce predominantly socially-directed vocalizations in circumstances where parents are trying to interact with infants (*Engaged*) and predominantly endogenous vocalizations when parents are interacting with another adult while the baby is in the room (*Independent*).

Materials and Methods

Approval for the longitudinal research that produced data for this study was obtained from the IRB of the University of Memphis. Families were recruited from child-birth education classes and by word of mouth to parents or prospective parents of newborn infants. Interested families completed a detailed informed consent indicating their interest and willingness to participate in a longitudinal study on infant sounds and parent-child interaction.

We selected six parent-infant dyads (3 male, 3 female infants) from the University of Memphis Origin of Language Laboratory's (OLL) archives of audiovisual recordings. The dyads had been recorded while engaged in naturalistic interactions and play. The three female infants were initially selected for coding in an earlier study on imitation (Long et al., 2016) which had utilized a coding methodology for judging illocutionary force similar to the one used in the present study. Three males were thereafter selected from the archives in order to balance the sample for gender. The selection was unbiased with regard to social vs. endogenous vocalization. All families lived in and around Memphis, Tennessee, and all but one infant were exposed to an English-only speaking environment (Infant 6 was exposed to English and Ukrainian at home). Parents were asked to speak English and no other language during the laboratory recordings. Criteria for inclusion of infant participants included a lack of impairments of hearing, vision, language, or other developmental disorders. Demographics and recording ages for each infant at each recording session are provided in Table 2.

Table 2. Infant demographics
All infants completed two recording sessions around 3, 6, and 10 months of age.

Infant	Gender	Birth order	Maternal education	Home language	Age of recordings (months; weeks)					
					1	2	3	4	5	6
1	F	1	PhD	English	3;2	3;2	6;0	6;3	9;3	9;3
2	M	2	BA	English	4;2	4;2	6;0	7;2	11;2	11;2
3	M	1	Some college	English	3;2	3;2	5;0	6;0	10;0	10;0
4	F	1	Some graduate school	English	3;0	3;0	5;0	6;0	10;1	10;1
5	M	3	Some college	English	3;2	3;2	6;0	6;3	9;3	9;3
6	F	1	PhD	English, Ukrainian	4;0	4;1	6;0	7;0	11;3	11;3
Nominal age of recording					**3 months**		**6 months**		**10 months**	

Laboratory Recordings

Two laboratory recordings were selected from each of the 6 infants at approximately 3, 6, and 10 months, for a total of 36 sessions. The average session length was 19 minutes (range: 12-22 minutes). During recordings, the parent-infant pairs occupied a studio designed as a child playroom with toys and books. Laboratory staff operated four or eight pan-tilt video cameras located in the corners of the recording studio from an adjacent control room—there were three such recording laboratories at varying stages of the research. In all the laboratories, two channels of video were selected at each moment in time with the goal of recording: 1) a full view of the interaction or potential interaction, including the infant and any potential interactors (i.e., parent or laboratory staff) with one camera and 2) a close view of the infant's face with the other camera. Both the parent and the infant wore high fidelity wireless microphones, with the infant microphone <10 cm from the infant's mouth. Detailed descriptive information regarding the recording equipment can be found in previous studies from this laboratory (Buder et al., 2010; Warlaumont et al., 2010).

In roughly counterbalanced orders across ages, parents were either instructed to interact with the infant (the expected *Engaged* circumstance) or with another adult while the baby was in the room (the expected *Independent* circumstance). Later at the same age (usually on the same day), the dyad was recorded in the other circumstance. Parents were asked to interact with the infant and/or laboratory staff in a naturalistic manner. During the expected *Engaged* circumstance, parents were encouraged to engage in face-to-face interaction with the infant but were not restricted from interaction with others if someone came into the room (e.g., to adjust cameras, to answer parent questions, etc.). Similarly, in the expected *Independent* circumstance, parents were encouraged to keep their attention and interactive focus on the laboratory interviewer but were not restricted from engaging with the infants if they appeared uncomfortable or if the infants were repeatedly bidding for attention. The freedom allowed in these naturalistic recordings resulted in variation in the actual circumstance with respect to the expected circumstance. Our analysis took account of social directivity of infant utterances in the actual circumstances only.

Coding for Engaged and Independent Circumstances

As indicated above, the recordings had been intended to be differentiated neatly as primarily corresponding to *Engaged* or *Independent* circumstances, but the infants often sought attention from the parents during sessions intended by protocol to be *Independent,* or adults would engage in conversation with a staff member during sessions intended to be *Engaged*. For this reason, we re-categorized segments of time within each session in terms of whether they were actually *Engaged* or *Independent*. Figure 3 exemplifies this re-categorization.

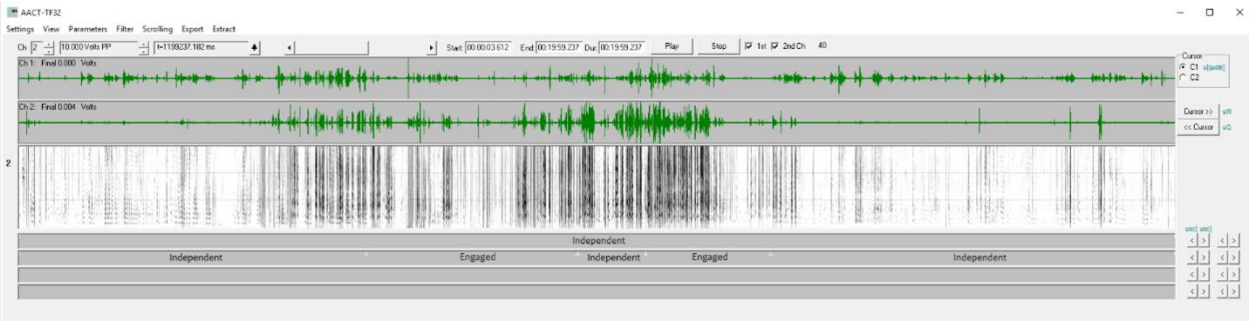

Figure 3. Visualization of re-categorizing circumstance
An example of one 20-minute recording (Infant 5 at 3 months) with the expected circumstance according to the protocol on line 1 of the coding field (below the spectrogram) and the re-categorization of actual circumstances on line 2. In this recording session, the parent was instructed to engage with the interviewer in accord with the Independent circumstance, but there were two substantial periods of time where the parent was actually directly engaged with the infant, and so those segments were re-coded as Engaged.

These re-categorized segments were used in the analysis of the role of circumstance in the infant utterances. Table 3 shows the re-categorized, actual circumstance durations for each infant and infant age. Appendix G provides a more detailed breakdown of expected and actual circumstance durations for each infant and infant age.

Table 3. Actual circumstance durations
Duration of actual circumstance segments Engaged (Engd) and Independent (Ind) for each infant at each age. Overall, there were longer periods of time in the Engaged circumstance than in the Independent circumstance. The minimum duration was 00:58, maximum duration 32:52, with an average duration of 19:06.

Mean age		3 months		6 months		10 months	
Infant	Gender	Engd	Ind	Engd	Ind	Engd	Ind
1	F	00:32:38	00:01:16	00:33:48	00:04:23	00:20:34	00:19:22
2	M	00:27:59	00:12:24	00:26:59	00:14:53	00:23:34	00:18:08
3	M	00:22:46	00:21:19	00:23:08	00:17:28	00:25:35	00:07:29
4	F	00:23:26	00:15:15	00:10:31	00:25:08	00:24:27	00:15:16
5	M	00:22:00	00:14:02	00:20:54	00:18:11	00:21:45	00:19:55
6	F	00:35:52	00:01:37	00:25:33	00:00:58	00:24:02	00:15:00

The amount of time pertaining to the actual circumstances that occurred during the recordings varied substantially, including two periods of time that included so few utterances (< 5) we did not include them in the analyses, as indicated in the total protophone counts of Table 4. This substantial variation in circumstance duration, along with the variability of actual ages

provided motivation for a statistical modeling approach that was robust and conservative with regard to such variations (see below).

Table 4. Protophone counts
Total counts of the number of protophones for the Engaged (Engd) and Independent (Ind) circumstances at each age for all infants. Cells marked with an asterisk (*) were excluded from analysis because they included fewer than 5 protophones.

Mean age		**3 months**		**6 months**		**10 months**	
Infant	**Gender**	**Engd**	**Ind**	**Engd**	**Ind**	**Engd**	**Ind**
1	F	446	4*	310	47	182	118
2	M	230	202	181	122	108	70
3	M	311	163	158	102	133	81
4	F	273	227	103	384	233	138
5	M	328	257	330	147	89	117
6	F	442	13	381	4*	116	107
Average		**338.33**	**144.33**	**243.83**	**134.33**	**143.5**	**105.17**

Coding of the Function of Infant Protophones

Coding for circumstance, illocutionary function, and gaze direction was completed within the Action Analysis Coding and Training software (AACT) (Delgado et al., 2010). This coding software has been used and discussed extensively in previous research from this laboratory (Jhang et al., 2017; Warlaumont et al., 2010; Yoo et al., 2018). The software affords frame-accurate coordination of video and audio, which is displayed in a special version of the TF32 software (Milenkovic, 2001). TF32 includes both flexible waveform and spectrographic displays. Coders can view and listen with a scrolling audio display where a cursor indicates the location of the audio at each moment of playback. The utterances to be coded in the present work were labeled for vocal type and bounded in time for onsets and offsets in AACT in prior studies (Oller et al., 2013). The AACT software allowed the coder to advance to each bounded utterance in turn for playback and coding in illocutionary force and gaze direction for the present study. The AACT software also allows users to export data that indicate whether an utterance occurred within an *Engaged* or *Independent* circumstance.

All infant protophones that had been previously bounded were also labeled for the present work in terms of *illocutionary force* (Austin, 1962; Oller et al., 2016; Searle, 1969) to indicate potentially communicative functions. Illocutionary force was originally defined by Austin as the social intention of a speech act, but has been extended in work in child development and animal communication to also encompass vocal acts produced with little or no social intention (Oller et al., 2013). In this extended usage, vocal play, for example, is treated as an illocutionary force. Another example: a fussy protophone, not directed toward anyone, can be treated as having the illocutionary force of complaint.

Pre-linguistic infants express varying illocutionary forces and varying emotional content (i.e., positive, neutral, and negative) in early protophones beginning at birth (Jhang & Oller, 2017; Oller et al., 2013) (see Appendix H). This fact indicates that infants have the capacity to produce a single protophone type with different illocutionary forces on different occasions, indicating they possess a vocal capability that is, of course, required of all words and sentences in mature language. Put another way, infant protophones can be used with varying communicative intentions, for example, to gain attention, to continue vocal interaction when engaged with a caregiver, or to make a request. The same vocalization types can also be produced for the infant's own purposes when not engaged in social interaction at all, e.g., when vocalizing toward an object or when simply exploring sound for its own sake.

The determination of whether a vocalization is social or endogenous requires considering a variety of factors. One is gaze direction during infant vocalization, but another is the extent to which infants may bid for attention vocally even when they are not in the same room with caregivers. Judging directivity of infant vocalizations also requires taking into account the relative timing of infant and caregiver utterances as well as the content of utterances of adults

who are present at the time of the recording, especially caregivers who presumably know a good deal about the capabilities of a particular infant. We make the assumption for this work that judgments about vocal directivity need to be made moment by moment, utterance by utterance, to account for the possibility that infants may engage and disengage in protoconversation. The judgments of the social or endogenous nature of infant protophones need to be made taking account of the broad context of events prior to and subsequent to each infant utterance, and factors such as timing, eye contact, perceived imitativeness, and meaningful responsivity must be allowed to yield intuitive judgments by the observer, where a balance among the factors provides the basis for the coding.

A coding scheme was created for making judgments on the illocutionary function of individual infant vocalizations in consideration of all of the above listed factors. *Social* protophones were labeled as such when, for example, the infant used them to initiate conversation, continue an ongoing interaction, imitate another person, or to complain or exult in a way that was directed to an adult as indicated by gaze, gestures, or other contextual factors. *Endogenous* protophones were identified as utterances infants produced for their own purposes; such events included vocal play, object-directed sounds, complaints and exultations not directed to others, or protophones with no clear illocutionary force. Brief descriptions of each code used for judgments of illocutionary function are provided in Table 5.

Table 5. Coding scheme for judgments of illocutionary function

Codes used for labeling illocutionary function of infant vocalizations. Contextual information such as gaze, body positioning, and timing was considered to make intuitive judgments on each infant utterance.

Endogenous vocalizations		**Social vocalizations**	
No Force	Produced without obvious exploratory or social intention	*Call/Initiate*	Call or bid for attention directed toward another person
Vocal Play	Not directed to a person or object but apparently playful	*Continue*	Maintenance of a turn taking sequence with another person with communicative intent
Object-Directed	Directed toward a toy or other object as indicated by body positioning, gaze, or gesture	*Imitation*	Matching of pitch or articulatory characteristics of another person's utterance while engaged in turn taking
Complaint	Distress vocalization not directed to another person	*Complaint-Directed*	Distress vocalization directed to another person
Exultation	Celebratory vocalization not directed to another person	*Exultation-Directed*	Celebratory vocalization directed to another person

Our coding is founded on the assumption that human observers are naturally able to judge the extent to which vocalizations at any age are intended as communicative acts—otherwise how would humans know when to respond or participate in vocal engagement? If some parents are poor at making such judgments, they are surely at a disadvantage in child rearing, because they don't know when their infants are communicating or not. It makes sense that natural selection has produced parents (and potential parents) that are capable of recognizing when infants are communicating intentionally and when not. Consequently, the coding process takes advantage of natural capabilities of human observers and gauges the extent of their reliability by comparing agreement among observers.

During illocutionary coding, both the primary coder and an independent reliability coder took a broad view of each utterance and its context of production. The coding was conducted by watching the entire recording session. Then the coder started at the beginning of each session and observed everything that happened up to the point of each infant utterance, and then coded with repeat observation. That is, each time a protophone was located, the judgment of illocution was

made based on the entire preceding context and the cursors could also be stretched so that, during repeated playbacks before coding for illocutionary force, the coder could, if necessary, see and hear the utterance plus a several-second context both before and after it repeatedly. If there was ambiguity about how to judge the possible social directivity of the utterance, the boundaries could be stretched further until the coder felt confident that no further stretching would improve the coding decision.

Coding for Gaze Direction of Infant Protophones

Gaze direction coding was conducted independently of the illocutionary coding for all protophones and was based on gaze direction only. For this coding, sound was turned off, and the coder determined whether at any time during each utterance, the infant looked toward another person. The time frame of playback for the period during which the protophones occurred was expanded through a special setting in AACT by 50 ms before and 50 ms after the actual utterance boundaries as indicated based on the original protophone coding. This expansion of time frame for viewing was deemed important because of the low frame rate of video recording (~30 ms per frame) and ensured that the entire period of the vocalization was available for visual judgment. Utterances could be played repeatedly this way. They were judged as "directed to a person" (during any portion of the utterance plus or minus 50 ms) or "not directed to a person" (during the same period). For utterances that included no good camera view of the infant (the infant sometimes turned away from the selected cameras and vocalized before new cameras could be selected) or for utterances where the infant's eyes were closed, the coder indicated "can't see" or "eyes closed," respectively. The gaze direction analysis excluded all such utterances. A brief description of each code used for judgments of gaze direction is provided in Table 6.

Table 6. Coding scheme for judgments of gaze direction
Codes used for labeling directivity of infant gaze during vocalization. Each infant utterance was also coded for gaze to provide a secondary analysis on social directivity of protophone production.

Directed Gaze	*Directed to Person*	Gaze clearly directed to another person's eyes or face
Gaze Not Directed	*Not Directed to Person*	Gaze clearly not directed toward another person
	To Toy	Gaze clearly directed toward a toy
	To Mirror	Gaze clearly directed into a mirror toward self or object in room and clearly not toward another person
Unclear Gaze	*Can't See*	Infant briefly outside of camera range; unable to make judgment
	Eyes Closed	Infant's eyes closed; gaze judgment not possible
	Unspecified	Gaze directed in the vicinity of person, unable to make a definitive judgment (e.g., too far away)

Coder Training and Coder Agreement

For the coding in the present study, both the primary coder and the agreement coder were trained in infant vocalizations and illocutionary coding by the last two authors in a training sequence that has been described in several prior publications (Oller et al., 2013; Oller et al., 2019; Yoo et al., 2018). In brief, the training included 1) a series of 5 lectures on vocal development and coding of early vocalization and interaction, 2) an interleaved set of corresponding coding exercises using recorded data like that to be encountered in the current research; 3) comparisons of the outcomes of those coding exercises with regard to outcomes for other coders, with special reference to coder agreement and agreement with gold standard coding by the last author, who has been engaged in vocal development research for more than 40 years (Oller et al., 1976); and 4) a certification process that resulted from reviews ensuring that coding results correlated highly with group coding and the gold standard coding and did not diverge from gold standard coding by more than 10% of mean values.

All the data of the present study were coded for illocutionary force (from which socially- and endogenous categories could be derived) by the first author, and approximately 30% of the

total data set was coded independently for illocutionary force by the agreement coder. An original coding of gaze direction had been done on three of the six infants by a previous team of coders for the paper previously cited (Oller et al., 2013). This completely independent prior coding on half of the data for the present study was available to offer an agreement check on the gaze coding done for the present paper.

Results

Protophone Usage Judged in Terms of Illocutionary Functions

A total of 6,657 infant protophones were labeled across all 36 recordings (6 infants x 3 ages x 2 sessions). The data account for all infant utterances that were judged to be non-vegetative (burp, hiccough) and not fixed signals (cry, laugh) across the 36 laboratory recording sessions. Utterances where either gaze or illocution could not be judged were eliminated. Two segments were eliminated from analysis because of a very low number of protophones for that infant at that age in that condition (specifically, Infant 1, *Independent* at 3 months and Infant 6, *Engaged* at 6 months, see Table 3 in Methods). Only 8 protophones occurred in these 2 segments. We also limited the analysis to include utterances that could be judged based on audio and video both for illocutionary force and for gaze direction. The final set included 6,388 protophones.

To determine if the usage of endogenous protophones exceeded that of social protophones, we used *t*-tests comparing percentages of endogenous protophones against 50%. To test for effects of Age (3 levels) and recording Circumstance (*Engaged* vs. *Independent*), a different approach was required. We selected a logistic regression model based on Generalized Estimating Equations (GEE). GEE analyses are a non-parametric alternative to generalized linear mixed models that accounts for within-subject covariance when estimating population-averaged

model parameters (Liang & Zeger, 1986). GEE is particularly appropriate for the data in question because of the unequal amounts of data in the two circumstances and the lack of precise age matching across infants. GEE provides a conservative but robust method for such cases.

Figure 4 displays the overall percentages of protophones produced by the six infants across the two broad illocutionary groupings of endogenous and social. Infants used significantly more endogenous protophones across the three ages than social ones, with about 75% of all protophones being endogenous. By *t*-tests of the percentage of endogenous protophones, it was found they significantly ($p < .001$) exceeded 50% at all three ages. We found no notable change in the predominance of the endogenous protophones across Age, and indeed the GEE revealed no significant difference in the percentage of social protophones across Age ($p = 0.48$). A subsequent GEE analysis was conducted with Age as a continuous variable and produced the same pattern, with more endogenous protophones than social ones ($p < .0001$) and no Age effect ($p = .69$).

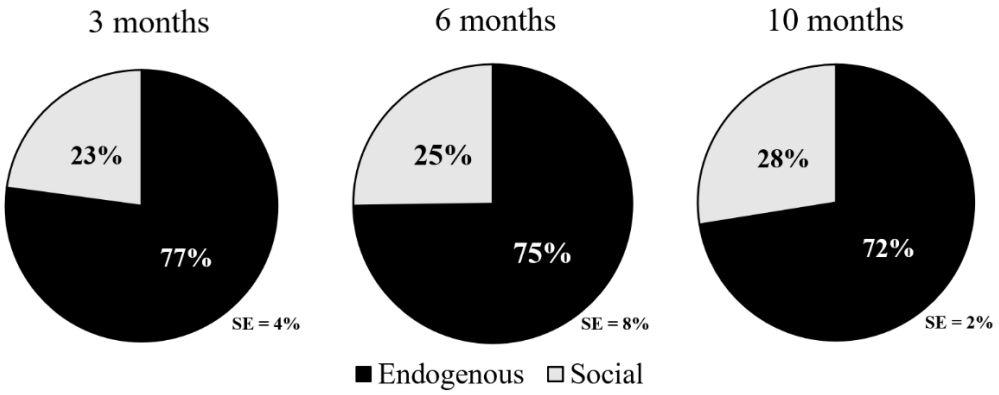

Figure 4. Social and endogenous infant protophones across 3 ages
Percentage of infant protophones that were judged to be endogenous (produced for the infants' own purposes) and social (overtly communicative) across all observations. Overall, infants primarily produced endogenous vocalizations (~75%), suggesting that the great majority of infant sounds are produced independent of social engagement in the first year. Furthermore, a non-significant main effect of Age is consistent with an interpretation of stable use of both social and endogenous protophones across the three ages.

Similarly, *t*-tests of the proportion of endogenous protophones in the two circumstances (*Engaged* vs. *Independent*) showed that endogenous protophones significantly exceeded 50% in both circumstances ($p < .001$). Based on the GEE for data presented in Figure 5, infants used significantly more endogenous protophones in the *Independent* circumstance than the *Engaged* circumstance ($p < .03$). A separate GEE analysis in which only main effects were considered revealed a stronger Circumstance effect ($p < .0001$). The fact that endogenous protophones outnumbered social ones in the *Engaged* circumstance contradicted our hypothesis and highlighted the predominance of endogenous infant vocalization. A separate GEE analysis of the data treating Age as a continuous variable yielded similar results. Specifically, significant differences were seen for overall proportions of protophones between circumstances ($p < .001$) and non-significant differences across Ages ($p = .982$).

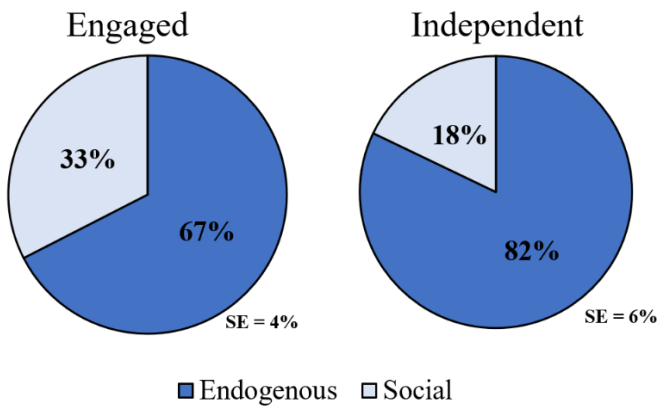

Figure 5. Social and endogenous infant protophones across two circumstances
Percentages of social and endogenous infant protophones across Engaged (parent and infant interacting) and Independent (parent and interviewer conversing while infant present in room) circumstances. Endogenous protophones predominated in both conditions.

The pattern of results revealed by the illocutionary coding was similar for both the primary coder and the reliability coder, with 79% point-to-point inter-rater agreement on 30% of the recordings that were coded independently by the two observers. For both coders, endogenous protophones predominated, and the reliability coder—who had no knowledge of the hypotheses

for this study—identified a slightly higher proportion of endogenous protophones (79.2%) than the primary coder (78.5%).

Protophone Usage Based on Gaze-Direction Judgments

As a check on the illocutionary coding, we considered an alternate, simpler way of gauging the function of infant protophones. The first author coded gaze direction during protophone production as being directed or not directed toward a person. Gaze judgments were made with sound off (video only) for all six infants.

Even though the function of protophones as determined by gaze-direction was not always the same as the function based on illocutionary judgments, the overall percentages of social protophones as determined by the two methods was very similar. That is, the great majority of infant protophones were judged to be produced with gaze directed somewhere other than towards any person in the room, just as the illocutionary judgments indicated the great majority of infant protophones to be endogenous. 72% of the infant protophones were deemed *not* to include person-directed gaze, while 75% were deemed endogenous by illocutionary coding.

In the earlier study mentioned above (Oller et al., 2013), 50% of the current sample had been coded for gaze direction, allowing for a robust analysis of independent inter-rater agreement. Inter-rater agreement on a point-to-point basis was 87% (of 3347 utterances). The results showed a strong predominance of protophones *not* being associated with gaze directed toward another person for both the earlier coders and the present one. Based on the same sample of utterances, the primary coder in this study found 64% of the utterances not to include person-directed gaze, while the previous (reliability) coder found 61% not to include person-directed gaze. These percentages represent only half the total sample (three of the six infants) and consisted heavily of samples from the *Engaged* circumstance; consequently, the percentages (64

and 61%) are lower than the 72% of utterances deemed not to include person-directed gaze for the whole sample as reported above.

Let us expand on why the gaze-direction and illocutionary coding methods do not yield exactly the same outcomes on the function of infant protophones. In the coding of illocutionary force, momentary gaze direction by the infant toward a person was sometimes not deemed to indicate the function of the vocalization. For example, a momentary glance directed to the parent occasionally occurred even though the infant appeared to be engaged in vocal play. There were also a number of cases where the coder deemed a protophone to be social in illocutionary coding, even though gaze direction toward a person was deemed absent. Such cases often corresponded to interactional sequences where the relative timing of utterances suggested the infant was engaged and directing the protophone to the parent, even though the infant was looking away.

Discussion

Overall, infants used about three times as many endogenous protophones as social ones. This predominance remained stable across the three ages. Even in the *Engaged* circumstance, where parents were trying to engage with their infants, endogenous protophones predominated, with twice as many judged to be endogenous as social. In the *Independent* circumstance, where parents were engaged in conversation with laboratory staff, the endogenous protophones predominated to a substantially greater extent, with four times as many endogenous as social.

The low rate of socially-directed vocalizations of infants in the first 10 months as reported here has required us to reorient our thinking about the functions of infant protophones. It seems important to draw attention to the fact that for all the sessions of recording reported on here the caregivers and infants were in the same room, and caregivers were aware that they were being recorded. The caregivers also knew the study was about vocal development, and it was

assumed they would endeavor to elicit infant vocalization and thus interact as much as possible. They often attended to infant vocalizations even in the designated *Independent* circumstances, sometimes responding to infant protophones with infant-directed speech (IDS), a pattern of caregiver responsivity that required some restructuring of our analysis to assign segments within sessions appropriately to the actual *Engaged* and *Independent* circumstances. Consequently, we presume parents tried to maximize their infants' socially-directed vocalization—and yet the rate was low.

Partly because the *Independent* circumstance resulted in a considerably larger predominance of the endogenous protophones than the *Engaged* circumstance, we presume that even more naturalistic recordings might produce an even greater predominance of endogenous protophones. That is, we suspect that the percentage of infant protophones that are socially directed in the natural environment of the home could be considerably *lower* than the values estimated here. This suspicion is supported by recent results where we compared the amount of IDS occurring in laboratory recordings for 12 infants (three of whom are among those represented in the present work) to the amount of IDS occurring in all-day LENA recordings (Zimmerman et al., 2009) conducted in the home with the very same infants at approximately the same ages across the first year of life (Oller et al., 2019). IDS was six times more frequent in the laboratory recordings than in randomly-selected five-minute samples from the all-day recordings when infants were awake. Thus, we reason that the percentage of endogenous protophones at home could be considerably higher than we have seen in the present work, since IDS is considerably lower. We plan to explore the rate of endogenous vocalization in all-day recordings in subsequent efforts. We also aim to study a larger sample of infants and to consider more differentiated circumstances of recording.

Our results contradict expectations that have often been apparent in the field of child development, where infant vocalizations are generally treated as responses to adult utterances or as attempts to engage adults in social interaction or to seek help from adults. Why has there been relatively low emphasis on exploratory or endogenous vocalization? It seems likely that the answer lies in the amount of attention given by caregivers to infant vocalizations that are directed toward them as opposed to those that are not. We assume parents and other caregivers notice and remember vocalizations that appear to be social in nature to a greater extent than endogenous ones, and perhaps developmental researchers are similarly influenced by the salience of infant sounds that are embedded in protoconversation. Furthermore, parents may attend to any unique type of spontaneously produced protophone—irrespective of the communicative intent—and adapt their behavior to promote continued production of that particular sound, creating the appearance of, or perhaps initiating engagement with the infant. Indeed, we have reported evidence suggesting caregivers pay greatest attention to salient vocal signals such as those occurring in imitation, even though vocal imitation is surprisingly rare in the first year (Long et al., 2019). Caregivers, and thus people in general, may be inclined to overestimate the proportion of salient vocal signals such as imitation or immediate responses in protoconversation since it seems likely these are the sounds to which parents attend the most. So when they render estimates, they tend to overstate the frequency of occurrence of the social ones. It is only with systematic counting of every vocalization occurring in recorded samples, as has been done in the present work, that it becomes possible to determine that the great majority of infant protophones are in fact directed to nobody.

The results strongly suggest, then, that babies vocalize predominantly for their own endogenous purposes, hundreds or even thousands of times daily—4-5 times per minute of

wakeful time based on randomly-sampled segments from all-day recordings at home (Oller, Griebel, et al., 2019). There is considerable evidence that not just in vocalization, but in other realms as well, babies are not passive learners and in fact regularly influence their own experiences (Bornstein, 2000). A fundamental question that requires answering based on the present work is: If protophones are not directed to caregivers, what is their purpose from a developmental or an evolutionary standpoint? What advantage could be associated with producing vocal sounds that are largely affectively neutral, produced most commonly in apparent comfort, but without social directivity (Jhang & Oller, 2017; Oller et al., 2013)?

One possibility is that infants may be learning the range of capabilities of their vocal system through sensorimotor exploration. We see evidence of this possibility when infants produce squeals for extended periods, repeatedly make small whisper sounds or raspberries, or babble the same syllables repeatedly to a toy. Of course it seems likely that endogenous and social vocalization both contribute to the development of the speech system (Piaget, 1952b; Stark, 1981). But importantly, the sounds infants use in endogenous vocal activity provide the raw vocal material that parents are able to use in engaging their infant in protoconversation.

Members of our research group and John L. Locke have argued elsewhere (Locke, 2006, 2009; Oller et al., 2016; Oller & Griebel, 2005) from an evolutionary-developmental (evo-devo) perspective (Carroll, 2005; Gottlieb, 2002; Kirschner & Gerhart, 2006; Müller & Newman, 2003) that high rates of endogenous infant vocalization and vocal play may constitute fitness signals. The idea is based on the fact that the human infant is altricial (born relatively helpless) and has a long road ahead of requiring caregiver assistance for survival—the need for such caregiving lasts literally twice as long as in our closest ape relatives (Locke & Bogin, 2006). Consequently, we have argued that the human infant experiences selection pressure on the

provision of fitness signals that could have the effect of eliciting long-term investment from caregivers, whose evolutionary goal can be portrayed as perpetuation of their own genes through grandchildren. From this point of view, caregivers should invest more in infants who seem healthy and tend to neglect infants who seem less healthy. We operate under the assumption that the production of comfortable vocalization can signal well-being and good health. This pattern of fitness signaling is hypothesized to have applied to the ancient hominin infant, who has been presumed in accord with the hominin "obstetrical dilemma" (Washburn, 1960), to have been more altricial than other apes as soon as humans were bipedal. In accord with the reasoning about bipedality—which proves surprisingly difficult to confirm in the fossil record (Gruss & Schmitt, 2015; Wells et al., 2012)—bipedality had narrowed the human pelvis and required the hominin infant to be born with a smaller head and brain and thus to be more altricial than other apes. While the roots of human vocal flexibility appear to lie in their value as fitness signals in a distant hominin past, modern human infants are not less altricial than their distant forebears, and consequently we reason that endogenous protophones continue to be under selection pressure as fitness signals in human infancy.

One might ask, if fitness signaling is the primary advantage of protophones, why do infants not endeavor to direct their protophones primarily toward potential caregivers?[1] Of course, some of the time they do, as indicated by our data. When they do not, the protophones may still be heard and noticed, if only semi-consciously by potential caregivers. A parent may

[1] An additional question is, are infants who produce more socially-directed sounds at a greater advantage? Because we know social interaction is necessary for language learning, it would be reasonable to assume that infants who produce more socially-directed sounds may be more likely to attract the attention of caregivers more frequently, and thus have greater exposure to experiences that support cognitive development. Future studies are needed to evaluate vocal directivity predictors for cognitive abilities.

hear comfortable infant protophones and draw the unspoken conclusion that the infant is well and needs no immediate attention. Regular events of noticing the infant's well-being may reinforce a caregiver's commitment to long-term investment precisely because it suggests that particular infant is healthy and thus likely to be a good investment for survival and reproduction. So it may pay for the human infant to produce protophones at prodigious rates in case someone might be listening.

The production of protophones in infancy at the beginning of the communicative split between ancient hominins and their ape relatives, perhaps millions of years ago, seems likely to have laid a foundation for a more extensive use of vocalization as a fitness signal later in life, for example, in mating or in alliance formation (Locke, 2009). And as the amount of protophone-like vocalization became more well-established in the hominin line, it surely provided a foundation for more elaborate uses of vocalization, ratcheting from simple fitness signaling toward more and more language-like uses (Oller et al., 2016).

Play is widely recognized as a theater for practice of the behaviors young mammals will need as they proceed through life (Bekoff & Byers, 1998; Lafreniere, 2011). But it is important to note that playful behavior can serve not only as practice, but also as a fitness signal for the altricial young of many species. Our suggestion is that protophones can be seen (in the substantial majority of cases) as playful indicators of well-being, but they would seem to contribute at the same time to a sort of preparation for the future in mating, in alliance formation, and ultimately (nowadays) in the development of language.

4. Social and Endogenous Motivations in the Emergence of Canonical Babbling: An Autism Risk Study (Long et al., in submission)

Abstract

There is a growing body of research emphasizing the role of intrinsic motivation and endogenous activity to support the development of cognitive systems alongside the well-established role of social interaction. The present study longitudinally evaluated canonical babbling across the second-half year of life, when canonical babbling becomes well-established. We compared segments rated as having high and low levels of turn taking and independent vocal play in 98 children at low and high risk for autism spectrum disorder. Segments were extracted from all-day home audio recordings to observe infants in naturalistic settings. Canonical babbling ratios (CBR) were determined based on human coding along with Likert-scale ratings on the level of turn taking and vocal play in each segment. We observed highly significant differences in CBRs between risk groups during high and low vocal play, but high and low levels of turn taking yielded a weaker effect. There were also interactions of CBR with age, risk, and vocal function variables. We conclude that social and endogenous/exploratory motivations may drive both high- and low-risk infant tendencies to produce their most speech-like vocalizations.

Introduction

Canonical babbling has been long established as a robust stage of prelinguistic vocal development occurring prior to the emergence of the first word, having been argued to constitute a necessary foundation for vocabulary development (Koopmans-van Beinum & van der Stelt, 1986; Oller, 2000; Stark, 1980). To our knowledge, there is no published research evaluating the role of exploratory motivation in infants' production of canonical babbling and no direct evaluation of the extent to which social engagement in vocal turn taking affects it. In the present research, we observed babbling in infants at low and high risk for autism in naturalistic contexts.

Segments extracted from all-day home audio recordings were rated for levels of infant turn taking and independent vocal play to measure the degree of social and non-social vocal activity (and thus, social and exploratory motivations, respectively). We examined these findings within an evolutionary developmental biology (evo-devo) framework (Bertossa, 2011; Carroll, 2005; Newman, 2000, 2012), in part to inform our understanding of how babbling may be used to signal developmental progress to caregivers (Locke, 2017; Oller & Griebel, 2005, 2008). Comparing differences between autism risk groups may help to elucidate exploratory tendencies and potential breakdowns in social motivation in autism, as well as providing clinically useful perspectives on the development of language foundations.

Canonical Babbling Development in Typical Development and Autism

Throughout the first half year of life, infants evidence an emerging capacity to control and coordinate the respiratory, phonatory, and articulatory mechanisms. Within the second half year, and rarely later than 10 months, infants begin canonical babbling (Oller, 1980; Stark, 1980), defined as the production of mature consonant-vowel syllables with well-formed transitions between the consonant- and vowel-like elements (e.g., [baba], [dada]). These syllables provide a basis for interaction and play with repeated and varied syllables, foundational for the production of first words (Oller, 2000). The onset of canonical babbling is known to be a robust predictor of typical speech development (Oller et al., 1998; Nathani et al., 2006), with delays observed in several disorders including deafness (Eilers & Oller, 1994; Oller & Eilers, 1988), Down syndrome (Lohmander et al., 2017; Lynch et al., 1995), Fragile X syndrome (Belardi et al., 2017), cerebral palsy (Levin, 1999; Nyman & Lohmander, 2018), and William syndrome (Masataka, 2001). Lang et al. (2019) reviewed the mixed evidence on canonical babbling onset in autism, summarized below.

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by deficits in social communication and restricted interests and repetitive behaviors (American Psychiatric Association, 2013). Diagnosis is common nowadays by 18-24 months of age (Zwaigenbaum et al., 2015). Symptoms in early infancy include reduced or absent dyadic interaction, social responsiveness, and joint attention (Kellerman et al., 2019; Mundy, 2017; Ozonoff et al., 2010), and there is some evidence suggesting prelinguistic vocal developmental anomalies (e.g., Sheinkopf et al. (2012)). Two studies previously analyzed *canonical babbling ratios* (CBRs) in infants with ASD. Patten et al. (2014) showed significantly lower ratios in children with autism at 9-12 months and 15-18 months compared to controls, and Paul et al. (2011) found lower ratios at 9 months in infants at high risk for autism compared to low-risk infants, but not in a 12-month group. Two retrospective video analysis studies also found mixed results when analyzing *canonical syllables per minute*. Werner et al. (2000) showed no differences between infants later diagnosed with autism relative to typically developing controls between 8-10 months but significant differences at 12 months in complex babbling rates, and Chericoni et al. (2016) found no differences between the two groups at ages 6-12 months. Two other studies observed *ages of onset* for the canonical babbling milestone in infants at low and high risk for autism. Iverson & Wozniak (2007) reported that high-risk infants had a wider range for age of onset for canonical babbling (5-18 months) compared to the low-risk group (5-9 months), but LeBarton & Iverson (2016) found 33/37 infants at high risk for autism reached the canonical babbling stage by 14 months, with a typical average mean age of onset (7.67 months). In a feasibility study analyzing *syllable complexity*, Pokorny et al. (2017) found that an equal number of neurotypical and autistic infants in each group (4/10) produced more complex types of utterances than single canonical syllables by 10 months.

Overall, there is a lack of conclusive evidence on canonical babbling developmental differences in children at risk for autism or later diagnosed with ASD. Inconsistent findings across studies may be attributed to the well-established variability in autism characteristics, the varying methodologies used to analyze babbling development, or the differing group types included (as pertaining to retrospective analysis of children diagnosed versus prospective risk studies). Additional research including larger sample sizes is also necessary to provide a smaller margin of error when comparing typically developing groups and groups with autism. In this study, we compare the emergence of canonical babbling for infants at low and high risk for autism using the largest sample size to date (98 infants) and with evaluation based on sampling from all-day recordings across the second half-year of life (483 total recordings).

The Social and Endogenous Nature of Infant Vocalizations

When evaluating the emergence of canonical babbling, there is reason to consider potential differences in social and endogenous motivations behind the production of these advanced vocal forms. Considerable research has evaluated the role of social interaction in infant vocal development and the emergence of language (Franklin et al., 2013; Gros-Louis et al., 2014; Hsu & Fogel, 2001; Iyer et al., 2016; Lee et al., 2018). Caregivers are known to elicit and maintain "protoconversations" (Bateson, 1975), supporting the emergence of mature vocal stages such as canonical syllables and words (Bråten, 1988; Golinkoff et al., 1992; Rochat et al., 1999). Experimental studies using the still-face paradigm have also shown effects of social interaction on infant volubility and vocalization types (Delgado et al., 2002; Franklin et al., 2013; Goldstein et al., 2009). It is important to note that this body of research has primarily examined the effects of parental interaction on infant behavior. To our knowledge, there is no published evidence

directly examining the relative roles of interaction and endogenous vocalization on infant vocal development, including canonical babbling.

Contradicting the perhaps implicit assumption that infant vocalizations are simply interactive, several researchers have recently emphasized the role of intrinsic motivation in the development of emotional and cognitive systems, including those related to vocal development (Davis & Panksepp, 2018; Moulin-Frier et al., 2014; Moulin-Frier & Oudeyer, 2013). Infants produce more speech-like vocalizations, or "protophones," (including both canonical and precanonical babbling) without person-directed gaze (both when alone and in the presence of caregivers) than they produce socially-directed sounds (Harold & Barlow, 2013; Oller et al., 2013). More recently, several authors of the present study found that approximately 75% of all infant protophones in laboratory recordings were endogenously produced (Long et al., 2020). We know that social interaction influences infant babbling, phonological learning, and complex language skills (Albert et al., 2018; Elmlinger et al., 2019; Goldstein et al., 2003; Goldstein & Schwade, 2008; Kuhl, 2007), but social and endogenous motivations for infant vocal activity require additional research to elucidate their relative roles. Thus, our research evaluates canonical babbling across segments with high and low levels of both infant turn taking and exploratory vocal play.

An Evolutionary-Developmental Perspective on the Role of Social Motivation in Canonical Babbling

We and others have hypothesized selection pressures on the production of endogenously produced protophones. Baby sounds can be seen as fitness signals selected to elicit long-term investment from caregivers, required across the lengthy period of relative helplessness, or altriciality, of infant humans (Locke, 2017; Long et al., 2020; Oller et al., 2016, 2019). In accord

with the fitness signaling hypothesis, the *quality* of infant vocalizations can be considered a salient and reliable signal of fitness. Following this reasoning, it might be seen as advantageous for infants to produce their most advanced vocal forms during periods of caregiver attention. Empirical evidence has been presented to show that caregivers are keenly aware of their infants' developmental capabilities, including in the vocal domain (Bodnarchuk & Eaton, 2004; Lyytinen et al., 1996; Oller et al., 2001). Higher rates of canonical syllables (as opposed to less advanced protophones) during social interaction than during periods of aloneness could suggest a social motivation for producing the more advanced protophones. If the idea is on target, we might conclude that canonical babbling was selected as a salient signal of developmental progress, especially during social interaction. Furthermore, a breakdown in the social motivation of infants as a result of a neurodevelopmental condition such as that seen in autism could potentially result in lower rates of canonical babbling during social interaction than in those of typically developing infants.

The *social motivation theory* (Chevallier et al., 2012) posits that reduced social attention in infancy leads to the social-cognition developmental differences observed in autism spectrum disorder. Additional research supports this notion, showing social information is less salient in individuals with autism (Chevallier et al., 2013; Schultz et al., 2000; Weeks & Hobson, 1987) and less intrinsically rewarding in individuals with autism compared to typical controls (Bottini, 2018; Gray et al., 2018; Scott-Van Zeeland et al., 2010; Sepeta et al., 2012). Reductions in social orienting can also affect language development (Baranek et al., 2013; Dawson et al., 2004; Su et al., 2020), a supposition supported by speculations predicting positive associations between social motivation and language emergence; these speculations have yielded, for example, the continuity hypothesis (Bruner, 1974), the speech attunement framework (Shriberg et al., 2011),

and the elicited bootstrapping hypothesis (Camarata & Yoder, 2002), which has been recently elaborated by Su et al. (2020). This body of research and theory highlights the importance of identifying early differences in social interaction in infants at risk for autism in order to provide support and intervention as early as possible.

Interestingly, there is limited research examining endogenously motivated vocal learning in infancy (Syal, 2011). Instead, the great majority of research has focused on parental activity rather than internal motivations of the infant as influencing vocal development. In a salient recent example of such research, Su and colleagues found early social motivations around 23 months predicted language skills 2 years later—specifically, higher performance on social motivation tasks was significantly correlated with functional language abilities (Su et al., 2020). Such literature is consistent with the expectation that reduced social attention and inclinations in early infancy may affect the infant's motivation to produce advanced vocal forms during interaction, and thus may yield reductions in vocal fitness signaling in infants with low social motivation. It is thus consistent with the social motivation theory and also with our evo-devo approach, to predict that infants with typically developing levels of social motivation will produce higher rates of canonical syllables during periods of high vocal interaction than infants with low social motivation. The present body of data offers the opportunity to evaluate this possibility during periods where caregivers and infants engage in high or low amounts of vocal turn taking, and while comparing canonical babbling rates of infants who are at low risk for autism (presumably with typical levels of social motivation) and infants who are at high risk for autism (presumably with lower levels of social motivation). In accord with the social motivation theory, we anticipate that low-risk infants will show higher rates of canonical babbling (with

respect to their own baselines) during periods of high turn taking, but that high risk infants will not show higher rates during high turn taking.

On the Role of Exploratory Vocal Play in Typical Development and Autism

We are influenced by the literature-based hypothesis that infants at low autism risk should be expected to produce more canonical syllables during social interaction, while high-risk infants should not be expected to do so, but recent evidence suggests a contrasting possibility. Research by Long et al. (2020) has shown that typically developing infants produce protophones (both canonical and precanonical) predominantly endogenously. Even in laboratory recordings, during periods when parents seek social interaction with infants, most protophones (~60%) appear not to be directed to parents, and this predominance of endogenous vocalization is even stronger (~80%) when parents are present with infants but not attempting to engage them. The results suggest that research on infant tendencies to vocalize at varying levels of advancement should compare circumstances showing high vocal turn taking with circumstances showing high endogenous vocal activity, which we shall refer to here as *vocal play* (Stark, 1980, 1981). Thus, we deem it important to examine not only social motivations for the production of canonical syllables but also intrinsic, exploratory motivations.

During vocal play infants explore sensorimotor aspects of the vocal apparatus and practice with various properties of sounds such as syllabic structure, amplitude, and pitch control. Play has been well established to be important throughout development. Piaget treated play as necessary for children to understand and learn about the world (Piaget, 1952). Vygotsky also viewed play as necessary for the development of cognitive systems and interpersonal relationships (Berk, 1994; Vygotsky, 1978). Panksepp and colleagues proposed play as a fundamental neurobehavioral process, motivated by a play “emotion”, distributed widely among

social animals (Davis & Panksepp, 2018; Panksepp, 2005; Panksepp et al., 1984; Panksepp & Biven, 2012). Stark described vocal play as highly variable, with infants producing sounds in new and repeated combinations, modifying patterns and features during bouts of independent infant vocal activity (Stark, 1980). Stark's description of vocal play evokes the notion that its occurrence can also be considered a sensorimotor exploration of the vocal mechanism necessary to learn and master speech production.

Although it appears that infants in general are endogenously motivated to produce protophones, the social motivation theory of autism hints at the intriguing possibility that infants with autism may be relatively more inclined to vocalize independently/endogenously than neurotypical infants. Further, the reasoning might be extended to suggest that the rate of canonical babbling would be relatively higher (with regard to their own baselines) for infants with autism than for typically developing infants. In the context of the present dataset, it might be predicted that infants at high risk for autism will produce relatively higher rates of canonical babbling during independent vocal play than infants at low risk for autism. In contrast, infants at low risk for autism would not be expected to show higher rates of canonical syllables during high vocal play.

These speculations are perhaps supported by the fact that children with autism have been shown to spend more time participating in isolated play with objects and to produce more repetition of physical actions in play compared to typically developing peers (Atlas, 1990; Naber et al., 2008; Sigman & Ungerer, 1984; Williams et al., 2001). These patterns suggest that as infants with autism begin to produce canonical syllables, they may be particularly interested in the physical, articulatory properties of these sounds—not unlike their often intense interest in the

physical characteristics of objects—and they may produce these sounds with greater repetition, perhaps in enjoyment of the self-stimulatory nature of the repetition.

Specific Aims and Hypotheses

The present research compares canonical babbling ratios (CBRs) of infants at low and high risk for autism during recorded segments with high and low levels of both turn taking and vocal play across three age ranges during the second half-year of life. We preliminarily analyzed for possible differences in CBR between infants of high and low socioeconomic status (SES) and found no significant differences; therefore, we do not report SES effects in the data below. Sex differences were evaluated in a recent study from our laboratory using the present dataset and no significant sex differences for CBR were found (Oller et al., 2020); therefore, we do not include sex as a variable in the present work. Findings from this study may inform our understanding of social and exploratory motivations in the emergence of advanced prelinguistic vocalizations in typical and atypical development. Furthermore, risk group differences could suggest early signs of social language impairments in the first year of life. The following are hypotheses to be evaluated:

Predicted Interactions

Our initial analyses conducted using Generalized Estimating Equations (GEE) addressed Turn Taking (TT) and Vocal Play (VP) separately. Consequently, one analysis included three variables: Age, Risk, and TT, and another: Age, Risk, and VP. Based on the social motivation theory of autism, we predicted interactions of:

1. Risk and TT: CBRs in low-risk (LR) infants will be higher during segments with high TT than low TT while high-risk (HR) infants will not show higher CBRs during high TT.

2. Risk and VP: CBRs in HR infants will be higher during segments with high than low VP while LR infants will not show higher CBRs during high than low VP.

We also examined the possible interaction between Risk and Age in a GEE analysis including only those two independent variables. We predicted:

3. Risk and Age: CBRs will increase to a greater extent in LR infants across the three ages than in HR infants.

Predicted Main Effects

For interpretive perspective, we also analyzed main effects for CBR in a final GEE including Age, Risk, TT, and VP. We predicted:

1. Age: Higher CBRs will occur at the later ages than earlier ages, highlighting infants' increasing ability to control the speech mechanism (Lee et al., 2018; Nathani et al., 2006; Oller, 2000).
2. Risk: Higher CBRs will occur in the LR group compared to the HR group, a prediction based on the predominant, albeit inconsistent findings of the existing literature (Lang et al., 2019).
3. TT: Higher CBRs will occur during segments with high TT compared to low TT.
4. VP: Higher CBRs will occur during segments with low VP compared to high VP.

Methods

The institutional review boards of the University of Memphis and Emory University approved the procedures used in this study. Families provided written consent prior to participation in this study.

Participants

As part of an NIH-funded Autism Center of Excellence conducted at the Marcus Autism Center in Atlanta, Georgia, 100 families of newborn infants were recruited via flyers, advertisements, social media and community referrals to participate in a longitudinal sibling study of development across the first three years of life. We analyzed data from 98 infants (two infants did not complete recordings at the ages studied in this paper). Infants were recruited as being either at high risk (HR, n=49) or low risk (LR, n=49) for autism. Infants were deemed HR if they had at least one older biological sibling with a confirmed autism diagnosis, and LR if they had no familial history of autism in 1st, 2nd, or 3rd degree relatives. Sex and socio-economic status (SES) measures[1] were balanced to the greatest extent possible in accord with known autism male-to-female ratios (Loomes et al., 2017) and SES make-up of participants living in the greater Atlanta, Georgia area who were willing and able to participate in a 3-year longitudinal study. Table 7 presents demographic information for the infants included in this study.

Table 7. Numbers of infants by Risk, Sex, and SES
Number of participating infants by risk status, sex, and socio-economic status (SES). *One infant's family did not report an SES level.

		High Risk	Low Risk	Total
Total		**49**	**49**	**98**
Sex	**Male**	34	30	64
	Female	15	19	34
SES*	**Low SES**	26	18	44
	High SES	22	31	53

Families were asked to complete audio recordings once a month between 1-36 months of age. This study used data collected between 6.5 and 13 months of age to represent the typical

[1] SES was measured using maternal education. Low/High-SES groups were based on a median split of maternal education in the entire cohort.

range of expected onset for and infant activity in canonical babbling. These data were grouped into three age ranges for analysis and labeled with reference to the approximate mean age within each group: 6.5-8.49 months (7.5 months), 8.5-10.49 months (9.5 months), and 10.5-13 months (12 months). It should be noted that the 12-month age group included a slightly smaller age range (1.5 months) compared to the 7.5- and 9.5-months age groups (2 months).

Audio Recordings

Audio recordings were completed using LENA recording devices (Gilkerson et al., 2017; Zimmerman et al., 2009). These devices are battery powered and secured inside the pocket of a special vest or clothing item with button clasps and can record up to 16 hours of audio per charge. LENA devices have a 16 kHz sampling rate for adequate play-back of audio for human coding judgements of recorded material.

Recording Procedures

Families completed all-day recordings once a month starting from the first month of life through the third year of life. Once a month, parents were provided with a LENA recording device and were supplied regularly with appropriately sized clothing for their child to wear throughout the day, as well as full instructions on how to carry out recordings. They were asked to turn on the recorder when their child woke up in the morning on the day of the recording and leave it running until the child went to sleep at night, in order to obtain a representative naturalistic recording of the child's whole day. They were asked to remove the recorder and leave it running nearby during bath times, sleep, and any situation where the recorder would press on the child's chest or cause discomfort. They were also allowed to pause the recording in any situation that they felt would violate their right to privacy or confidentiality. Recordings were scheduled for the same calendar day each month, as far as possible, to rotate weekdays. The

device was returned to the research project staff at the Marcus Autism Center each month following recording days for data processing. Each family completed ~5 total recordings (range: 1-7) across the ages studied, with an average recording time of approximately 11 hours per day.

Coding Procedures

Twenty-one 5-minute segments were randomly extracted from each recording and coded in real-time for infant utterance counts by 16 trained graduate student coders[2] at the Origin of Language Laboratory (OLL). OLL staff were blinded to all diagnostic and demographic information associated with each infant recording throughout the coding process. From these 21, eight segments with the highest infant vocalization volubility and a range of infant-directed speech[3] were selected for further analysis from each recording, totaling 3799 segments. Fifteen of these segments were later excluded on the basis of having no infant vocalizations; therefore, final analyses were completed on a total of 3784 segments.

Canonical Babbling Ratios as a Measure of Advanced Prelinguistic Vocal Forms

In a second pass of coding, the 8 selected segments were coded in real-time for infant canonical and non-canonical syllable counts. Listeners identified a total of 30,263 canonical syllables, and 233,877 noncanonical syllables across all segments. To measure the emergence of advanced vocal forms, a *canonical babbling ratio* (CBR) was calculated as the total number of

[2] Graduate student coders were trained to differentiate canonical and non-canonical syllables during real-time coding and to rate the extent to which infants produced socially interactive (TT) and endogenous (VP) vocalizations during completion of the questionnaire that was filled out at the end of coding of each 5-minute segment.

[3] The amount of infant-directed speech (IDS) was rated using the questionnaire that followed each of the 21 segments coded in the first coding pass. The questionnaire was also used to indicate environmental contextual factors for each segment, including audibility, other-person activity level, and aloneness of the infant. As with the second coding pass, each questionnaire item required a 5-point Likert-scale response to the relevant question, e.g., for IDS, "*How often did someone talk to the infant?*"

canonical syllables divided by the total number of syllables in each segment. Means and standard deviations of CBRs were calculated for each infant at each age. These data were then averaged within each age (7.5, 9.5, and 12 months) and risk group (HR and LR). Occasionally, families did not complete a monthly recording, and for those cases there was no data at the infant's age to include in the analysis. If there were multiple recordings per age and infant (occasionally two recordings were completed at a single age), the means and SDs of these recordings were averaged for analysis.

Turn Taking and Vocal Play as Measures of Infant Vocal Function

Following syllable coding of each 5-minute segment, coders answered a 17-item questionnaire regarding how often infants used vocalizations for various functions based on the audible context of the infant's environment in each segment. See Long et al. (2020) for theoretical perspectives on making intuitive judgments of infant vocal functions. We used two items from the questionnaire to measure frequencies of naturalistic infant vocalizations that were judged to be inherently social and exploratory within each segment. Specifically:

1. Turn Taking (TT): Were any of the infant's protophones used in vocal turn taking with another speaker?
2. Vocal Play (VP): Were any of the infant's protophones purely vocal play or vocal exploration?

Coders were instructed to respond to each question using a Likert Scale which aligned to the following rating designations: *1 = Never, 2 = Less than half the time, 3 = About half the time, 4 = More than half the time, 5 = Close to the whole time*. For example, a TT rating of 5 was applied to segments where a caregiver was clearly speaking to the infant, and the infant was vocalizing in an apparent back and forth vocal interaction for essentially the whole segment.

Segments with a VP rating of 5 would indicate the listener perceived the vast majority of infant vocalizations as playful and exploratory and not directed to another person in any way. TT and VP were not considered opposing vocal functions; in other words, a TT rating of 5 would not necessitate a VP rating of 1. In segments with very high infant vocal activity containing both interactive and non-interactive information, it is conceivable that a segment could be rated as having high TT (5) and high VP (5). Conversely, a segment with low infant vocal activity and limited interaction with the parent would have low TT and VP ratings.

Ratings for TT and VP were dissimilarly distributed across the Likert-scale range, as shown in Table 8. In order to compare levels of TT and VP with maximally similar numbers of segments at two levels in both cases, we split TT ratings into "No Turn Taking" (Rating of 1) vs. "Any Turn Taking" (Ratings 2-5), and VP ratings into "Low Vocal Play" (Ratings 1-3) vs "High Vocal Play" (Ratings 4-5) levels. Even with this procedure, the TT split yielded a dramatic imbalance, with more than 80% of all segments pertaining to the No TT grouping. On the other hand, VP was very common, with only 8% rated as having no VP, and 55% rated as having VP occurring either "more than half the time" or "close to the whole time."

Table 8. Frequency distribution of segments for TT and VP

Following the coding of infant syllables in 3784 segments, coders rated each 5-minute segment on the frequency of vocal turn taking (TT) and Vocal Play (VP) for infants at low risk (LR) and high risk (HR) for autism. The distribution of segments along the rating scale for TT and VP was similar for both risk groups. Ratings for each variable were combined into two levels for maximally similar numbers within each category: "No TT" (TT rating: 1) vs "Any TT," (TT: 2-5) and "Low VP" (VP: 1-3) vs "High VP" (VP: 4-5).

Likert scale rating	Interpretation (Level of occurrence)	TT Level	TT count		VP Level	VP count	
			HR	LR		HR	LR
1	Never	No TT	1564	1482	Low VP	164	147
2	Less than half the time	Any TT	295	312		307	254
3	About half the time		42	55		431	383
4	More than half the time		18	14	High VP	506	526
5	Close to the whole time		2	0		513	553

Coder Agreement

Inter-rater agreement was examined for CBRs, TT level, and VP level using a secondary LENA recording dataset coded by 7 of the same graduate student coders following the coding protocol used in this study. The 5-minute segments that had already been coded—each by one of the 7 individuals—came from a set of over 1000 such segments randomly selected from the all-day recordings of eight infants at each of six ages across the first year of life. >380 of these segments had been coded in the very same way as in the present study, with determination of CBR, TT, and VP. A subset of 212 of these segments was semi-randomly selected to be assigned for a second pass of agreement coding, where the agreement coder would always be a different individual from the one who had provided the original coding. The number 212 was based on available coder time and the desire for a large enough sample to yield trustworthy agreement data.

Every one of the 7 agreement coders was assigned to at least 5 segments that had originally been coded by each of the other 6 coders. In addition, all agreement coders were assigned to at least 5 segments from each of the 8 infants. Finally, all the ages of infants were included for assignments to each of the agreement coders for at least 5 segments. The agreement coding was conducted blind, in the sense that no coder knew who had originally coded the segments assigned to them in the agreement phase, nor were they supplied with information about age or identity of the infants.

The agreement coding for canonical babbling ratios revealed high agreement for both the entire set, with ages ranging across the entire first year (r = .89), and for the subset that pertained only to the second half year (r = .87), a time period during which CBR varies substantially above 0 across the entire range of ages. Both the questionnaire items yielded far better than chance

levels of agreement on the Likert-scale judgments categorized binarily as in the present work (No TT = 1, Any TT = 2-5; Low VP = 1-3, High VP = 4-5) based on Chi square analysis ($p < .001$). For VP there was agreement on 66% of pairings, while for TT there was agreement on 87%, with only fair agreement on kappa (TT = .40, VP = .33). This level of agreement should offer little surprise, given the subjective nature of the judgments. We have been surprised, however, by the power to significantly predict CBR that these blunt measures offer, as will be seen below.

Statistical Approach

We used Generalized Estimating Equations (GEE) implemented in R to analyze main effects and interactions of Risk, Age, and TT and VP on infant CBRs and also tested independently for main effects of all four independent variables. GEE analyses are an advanced form of modeling providing a non-parametric alternative to generalized linear mixed models for estimating within-subject covariance and population-averaged model parameters (Liang & Zeger, 1986). GEE has advantages over other mixed models frameworks especially in cases where data across conditions and from participants are intercorrelated and where numbers of observations per participant or condition varies. Another advantage is that the GEE approach requires no normality assumption. A GEE analysis is appropriate here because this is a longitudinal dataset with an unequal number of observations on infants, number of recordings per Age and Risk group, and number of observations of TT and VP ratings within each level.

Results

We ran three GEE models evaluating interactions and main effects for 1) TT, Age, and Risk, 2) VP, Age, and Risk, 3) Age and Risk, and we ran a fourth GEE model on main effects only for 4) Age, Risk, TT, and VP.

Turn Taking, Age, and Risk

1. Predicted interaction of Risk and TT: Based on predictions derived from the social motivation theory, we predicted higher *CBRs in low-risk (LR) infants during segments with high turn taking but no such pattern in high-risk (HR) infants*. However, the results (Table 3) did not confirm the hypothesis (p = .144). The mean CBR for HR and LR infants was quite similar for segments with no TT, but somewhat (though not significantly) higher in LR infants for segments with any amount of TT.

In the full GEE model, we found no main effect of TT, that is, no significant difference in CBRs between segments rated as having No vs Any TT (p = .347). Differences in CBRs between Risk groups were also non-significant (p = .111), with somewhat higher CBRs in the LR group. In the same model, the main effect for CBR from 7.5 to 9.5 months of age was highly significant (p < .001, b = .06), but differences from 9.5 to 12 months were not (p = .121), reflecting the fact that CBRs went up more from 7.5 to 9.5 months than they did from 9.5 to 12 months.

Table 9. Turn Taking, Age, and Risk interaction model
Full interaction GEE model for CBR with Age group (7.5 to 9.5, and 9.5 to 12 mo.), Risk Group (HR vs LR), and Turn Taking as a factor (No TT vs Any TT).

Variable	Effect size (*b*)	SE	*p*
TT (No vs Any)	0.02	0.02	0.347
Risk (LR vs HR)	-0.02	0.01	0.111
Age (7.5 to 9.5 mo.)	0.06	0.01	< .001
Age (9.5 to 12 mo.)	0.02	0.02	0.121
TT * Risk	0.04	0.03	0.144
TT * Age (7.5 to 9.5 mo.)	0.01	0.02	0.812
TT * Age (9.5 to 12 mo.)	0.02	0.02	0.426
Risk * Age (7.5 to 9.5 mo.)	0.05	0.02	0.004
Risk * Age (9.5 to 12 mo.)	0.03	0.02	0.075
TT * Risk * Age (7.5 to 9.5 mo.)	-0.04	0.04	0.368
TT * Risk * Age (9.5 to 12 mo.)	-0.05	0.03	0.136

The results did not show significant two-way interactions between TT and either of the Age group comparisons: 7.5 to 9.5 months (p = .812) or 9.5 to 12 months (p = .426). There was,

however, a significant interaction between Risk and Age for 7.5 to 9.5 months (p = .004, b = .05); CBRs increased in HR infants to a greater extent between 7.5 and 9.5 months compared to LR infants across these two ages. This difference was reversed from 9.5 to 12 months such that LR infants (p = .075) showed a greater increase than HR infants in that age interval, an interaction that approached statistical significance. No significant three-way interactions were observed in this model. Figure 6 provides graphic illustration of the results presented in the full model for Age, Risk, and TT level.

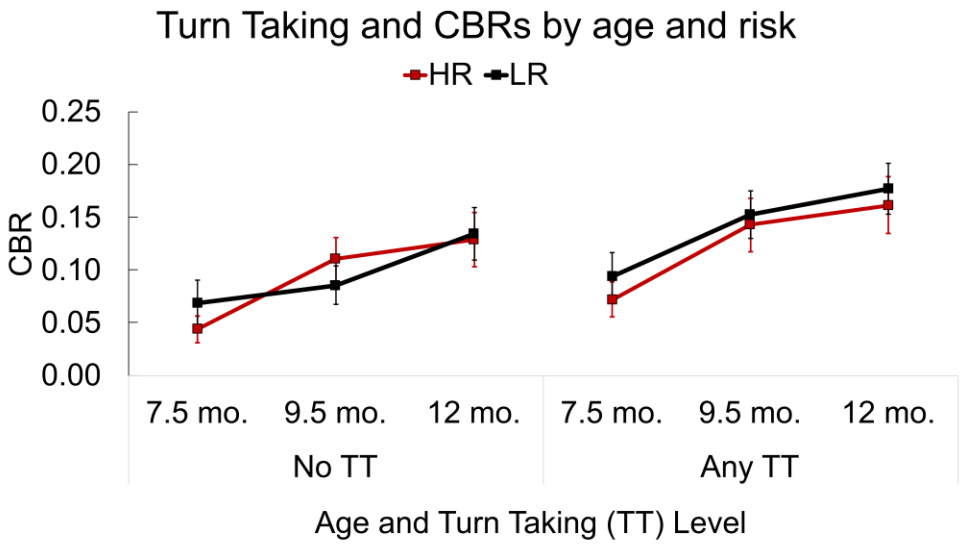

Figure 6. Canonical babbling by Age, Risk, and Turn Taking level
Canonical babbling ratios of infants at high risk (HR) and low risk (LR) for autism during segments with no vs any turn taking (TT) across three age ranges, 6.5-8.49 (7.5 months), 8.5-10.49 (9.5 months), and 10.5-13 (12 months) months. CBR was significantly higher from 7.5 to 9.5 months ($p < .001$, $b = .06$), and there was a significant two-way interaction of Risk and Age again between 7.5 and 9.5 months ($p = .004$, $b = .05$). There were no significant interactions including TT as a variable, including the three-way interactions of Age, Risk, and TT level. The values presented in the figure were computed from the raw data with means and SEs weighted for the number of infants who contributed data in each Risk group at each Age.

Vocal Play, Age, and Risk

In the full GEE model for Age, Risk, and VP (Table 10) we found several significant effects not observed in the model for Age, Risk, and TT.

Table 10. Vocal Play, Age, and Risk Interaction Model
Full interaction GEE model for CBR with Age (7.5-, 9.5-, and 12 mo.), Risk (HR vs LR), and Vocal Play (Low VP vs High VP).

Variable	Effect size (*b*)	SE	*p*
VP (Low vs High)	0.09	0.01	< .001
Risk (LR vs HR)	0.00	0.02	0.800
Age (7.5 to 9.5 mo.)	0.03	0.01	0.023
Age (9.5 to 12 mo.)	0.05	0.02	0.001
VP * Risk	-0.03	0.02	0.021
VP * Age (7.5 to 9.5 mo.)	-0.06	0.02	0.001
VP * Age (9.5 to 12 mo.)	-0.04	0.02	0.059
Risk * Age (7.5 to 9.5 mo.)	0.02	0.02	0.426
Risk * Age (9.5 to 12 mo.)	-0.01	0.02	0.679
VP * Risk * Age (7.5 to 9.5 mo.)	0.06	0.03	0.063
VP * Risk * Age (9.5 to 12 mo.)	0.06	0.03	0.039

2. *Predicted interaction of Risk and VP*: Based on predictions derived from the social motivation theory, we predicted *an increase in CBRs in HR infants from segments with low to high VP, and a lesser increase or no increase from low to high VP for LR infants*. There was indeed a significant interaction between VP level and Risk group (p = .021, b = -.03), but the direction of the effect was the opposite of that predicted. CBRs of LR infants increased to a greater extent from low to high VP than CBRs of HR infants. Based on calculations for Figure 7, CBRs at low VP were comparable (HR = .079, LR = .080), while those at high VP differed more, favoring the LR group (HR = .119, LR = .124).

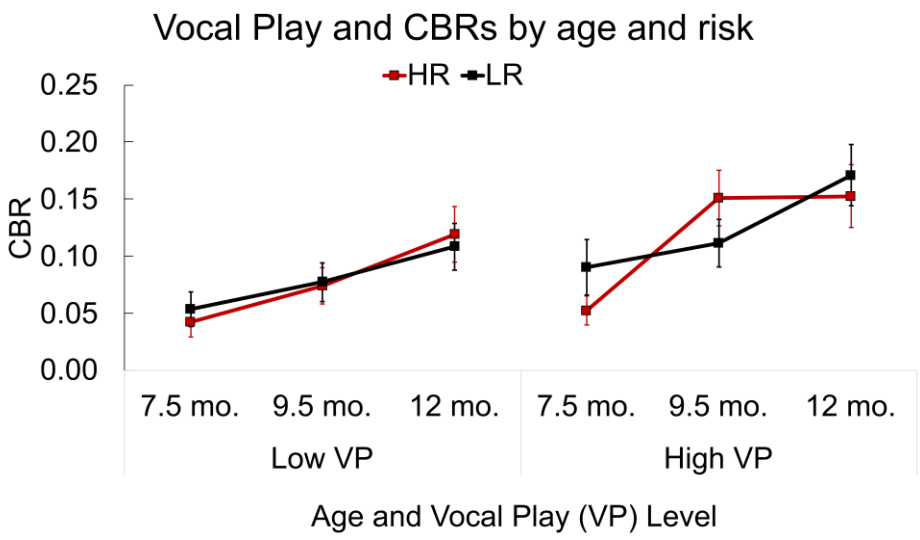

Figure 7. Canonical babbling by Age, Risk, and Vocal Play
Canonical babbling ratios (CBRs) of infants at high risk (HR) and low risk (LR) for autism in segments with low vs high vocal play (VP) across three age ranges, 6.5-8.49 (7.5 months), 8.5-10.49 (9.5 months), and 10.5-13 (12 months) months. In this model, there was a significant effect of Age for both 7.5 to 9.5 months (p = .023, b = .03) and 9.5 to 12 months (p = .001, b = .05). A significant interaction occurred between Risk and VP level (p = .021, b = -.03) and Age and VP level at 7.5 to 9.5 months (p < .001, b = -.06), with the interaction approaching significance for ages 9.5 to 12 months (p = .059). The three-way interaction among VP level, Age, and Risk was significant for ages 9.5 to 12 months (p = .039, b = .06), and approached significance for 7.5 to 9.5 months (p = .063). As in the case of Figure 1, the values presented here were computed from the raw data with means and SEs weighted for the number of infants who contributed data in each Risk group at each Age. Standard error (SE) bars are shown.

There was a highly significant main effect of VP, corresponding to a higher overall mean CBR produced by all infants during high VP compared to low VP ($p < .001$, $b = .09$). As with the full TT model, we observed no significant difference between Risk groups in the full VP model. There was, however, a significant effect of Age at both levels in the full VP model, with CBRs significantly increasing between ages 7.5 and 9.5 months, ($p = .023$, $b = .03$) and between ages 9.5 and 12 months, ($p = .001$, $b = .05$).

There was a significant two-way interaction for CBR between VP level and Age for 7.5 to 9.5 months ($p < .001$, $b = -.06$); this interaction reflects the fact that CBRs differed more between high VP and low VP at 9.5 than at 7.5 months. The difference between VP level and Age for 9.5 to 12 months approached significance ($p = .059$), and the effect was in the opposite

direction, namely CBRs differed less for high VP vs low VP at 12 than at 9.5 months. There were no differences between Risk and Age at either age comparison.

There was a significant three-way interaction between VP level, Risk, and Age for ages 9.5 and 12 months (p = .039, b = .06). The three-way interaction for Risk, VP level, and Age at 7.5 and 9.5 months approached significance (p = .063). Figure 7 provides a graphic display of the effects found with the second model and helps illustrate the nature of the three-way interactions. The data from segments rated as having high VP (right-hand panel) suggest a tendency of CBR to grow rapidly from 7.5 to 9.5 months in the HR infants, but to grow much less rapidly in the LR infants. The opposite growth pattern (LR more rapid, HR less rapid) is seen from 9.5 to 12 months. No such differentiation is observable in the left panel. Thus, the data suggest the LR and HR infants show very different patterns of growth in CBR with age, but only in cases of high VP.

Age and Risk

3. Based on the preponderance of prior research in autism, we predicted that *CBRs of LR infants would increase to a greater extent across the three ages than CBRs of HR infants*. The results did not conform simply to the prediction. In fact CBRs for HR infants rose *more* in the first age interval (from 7.5 to 9.5 months, ~.067 CBR units) than for LR infants (~.015), while they rose *less* in the second interval for HR infants (~.010) than for LR infants (~.065). These patterns corresponded to a significant interaction of Risk by Age at the first interval (7.5 to 9.5 months, p = .017, b = .04), but a non-significant interaction of Risk by Age at the second interval (9.5 to 12 months, p = .192).

Table 11 presents the full GEE model comparing Age and Risk groups. There was a significant main effect for both Age intervals (7.5 to 9.5 months, p < .001, b = .06; 9.5 to 12

months, $p = .047$, $b = .03$), suggesting an overall increase in CBRs over time, as expected. As in the prior models, there was no significant difference between Risk groups ($p = .319$).

Table 11. Age and Risk interaction model
GEE interaction model for CBR with Age (7.5, 9.5, and 12 mo.) and Risk (HR and LR) only.

Variable	Effect size (*b*)	SE	*p*
Age 7.5 to 9.5 mo.	0.06	0.01	< .001
Age 9.5 to 12 mo.	0.03	0.01	0.047
Risk	-0.02	0.02	0.319
Risk * Age (7.5 – 9.5 mo.)	0.04	0.02	0.017
Risk * Age (9.5 – 12 mo.)	0.02	0.02	0.192

Figure 8 illustrates these data, showing CBRs of LR infants increased only slightly in the first age interval and a much larger increase in the second interval. Conversely, CBRs in the HR group increased much more in the first interval than in the second. Comparing this interaction with the data in Figures 6 and 7 offer perspective. In Figure 6 (TT model), Risk and Age interacted such that the greater growth of CBR for HR infants in the first age interval applied primarily to the circumstance of No TT, although the three-way interactions corresponding to this observation were not significant. In Figure 7 (VP model), Risk and Age interacted such that the greater growth of CBR for HR infants in the first age interval applied primarily to the circumstance of high VP, and the three-way interactions corresponding to this observation were significant for the first interval and approached significance for the second.

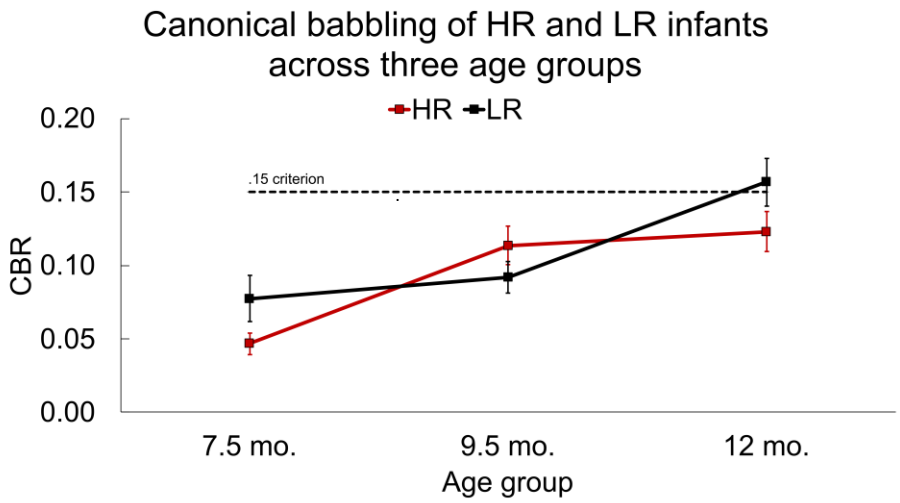

Figure 8. Canonical babbling ratios by Age and Risk
Canonical babbling ratios of infants at high risk (HR) and low risk (LR) for autism across three age ranges, 6.5-8.49 (7.5 mo.), 8.5-10.49 (9.5 mo.), and 10.5-13 (12 mo.). Overall, we found a significant interaction of Risk by Age for the first interval (7.5 to 9.5 months, p = .017), with CBRs rising much faster for HR infants than LR infants. The pattern was reversed, but not significantly in the second interval. Standard error (SE) bars shown.

A comment on the magnitude of the CBRs reported here seems warranted. The present data are based on all-day recordings sampled randomly; the CBRs are considerably lower than in prior reports based largely on short recordings usually conducted in laboratories and often selected for high infant volubility and/or interactivity. The Figure displays the criterion level of CBR that has sometimes been suggested to determine whether an infant is in the canonical stage based on a recorded sample (Lewedag, 1995; Oller, 2000; Oller et al., 2001; Patten et al., 2014). The mean CBR reached this .15 criterion for the LR infants only, and they reached it at 12 months only. The data suggest that the criterion level CBR for onset of the canonical stage should be considerably lower for all-day recordings sampled randomly than for laboratory recordings.

Main Effects

In a separate GEE model analyzing main effects only (Table 12), we found a significant effect of Age at both intervals (7.5 to 9.5 months, $p < .001$, $b = .04$; 9.5 to 12 months, $p < .001$,

b = .04), evidencing a strong and near linear increase of CBRs over time for data amalgamated across the Risk groups and independent of TT and VP. There was also a significant effect for both TT ($p < .001$, $b = .04$) and VP ($p < .001$, $b = .06$). The effect sizes, reflected in the *b* values from the GEE analysis, can be placed in perspective by considering that TT had an effect roughly of the same magnitude as 2-3 months of growth in CBR, and that VP had an even larger effect.

Table 12. Main effects for Age, Risk, TT, and VP

Main effects model for Age (7.5, 9.5, and 12 months), Risk (LR and HR), Turn Taking (TT) level (No TT vs Any TT), and Vocal Play (VP) level (Low VP vs High VP).

Variable	Effect size (*b*)	SE	*p*
Age 7.5 to 9.5 mo.	0.04	0.01	< .001
Age 9.5 to 12 mo.	0.04	0.01	< .001
Risk	0.004	0.01	0.742
TT	0.04	0.01	< .001
VP	0.06	0.01	< .001

The magnitude of the significant effects by Cohen's *d*, computed from the raw data—with means and SEs weighted for the number of infants who contributed data in each Risk group at each Age—was 0.29 (small) for both TT and VP. The Age effect size was 0.36 (small) for the first interval, 0.21 (small) for the second, and 0.55 (medium) for a comparison of 7.5 months with 12 months. There was no main effect of Risk ($p = .742$). Figure 9 displays these main effects, including significantly higher CBRs during both any TT and high VP compared to periods of no TT and low VP, respectively.

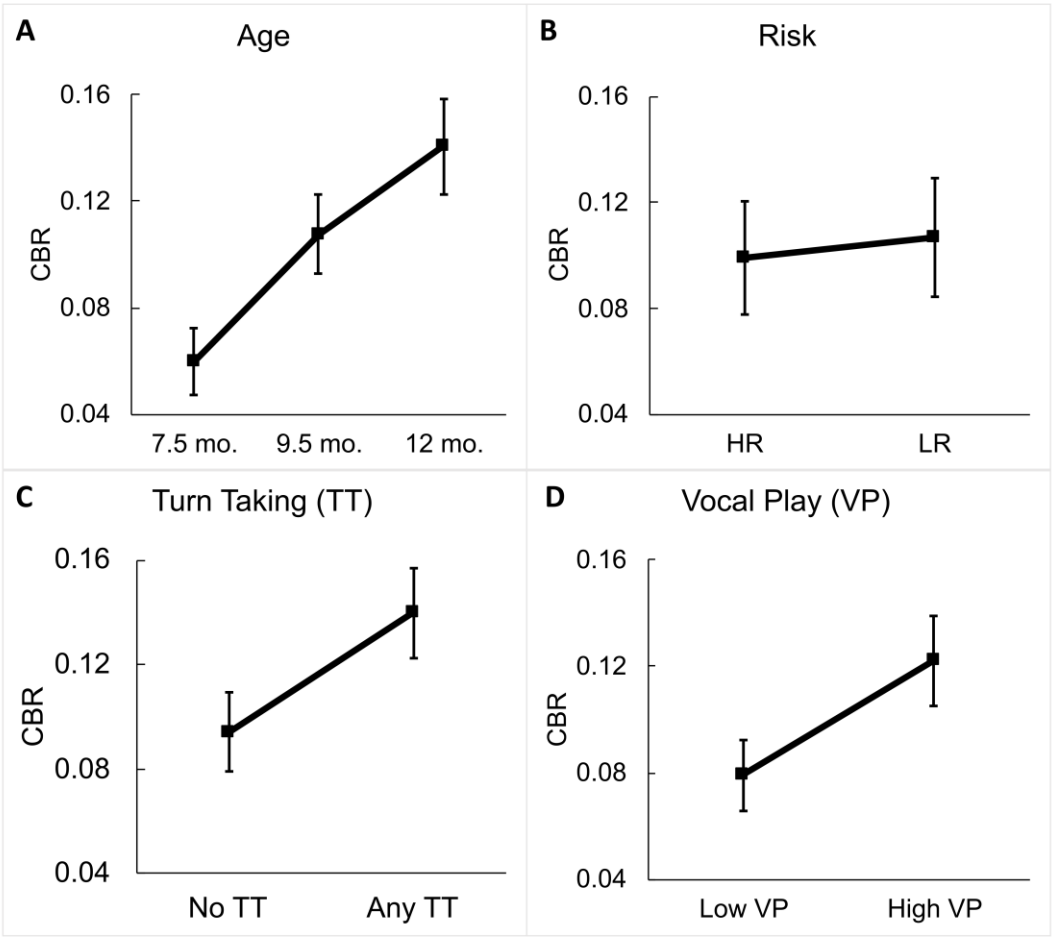

Figure 9. Main effects for Age, Risk, Turn Taking, and Vocal Play
Figure 9A illustrates the significant main effects of Age between 7.5 and 9.5 months ($p < .001$, $b = .04$) and 9.5 and 12 months ($p < .001$, $b = .04$). 9B shows the non-significant main effect of Risk group ($p = .742$). 9C presents the significant main effect of Turn Taking, with higher CBRs during segments rated as having any TT compared to those rated as having no TT ($p < .001$, $b = .04$). Finally, 9D shows the significant main effect of Vocal Play, with higher CBRs present during segments with high VP compared to segments with low VP ($p < .001$, $b = .06$). Standard error (SE) bars are shown.

Discussion

The present work evaluated the emergence of canonical babbling by comparing canonical babbling ratios (CBRs) in 98 infants either at low or high risk for autism across 3784 five-minute segments, selected from all-day recordings in the infants homes across the second half-year of life. The segments were coded by a team of trained listeners, who determined both CBRs and frequencies of vocal turn taking (TT) as well as vocal play (VP) in each segment. We addressed

these data with expectations derived in part from the social motivation theory of autism, assuming that infants at high risk (HR) for autism may show lower social motivation than infants at low risk (LR). We also considered the data in light of evo-devo, a biological perspective in which it has been posited that early language development is driven by interplay between social motivation (presumably reflected in infant interest in caregiver vocalizations and in protoconversation) and an endogenous inclination in infants to produce copious amounts of vocalization, one that appears to have been naturally selected as a signal of fitness. These theoretical views led us to propose ways that risk for autism might play an important role in the emergence of canonical babbling, the sort of infant vocalization that is required in order for word learning to be launched in earnest, since words are overwhelmingly composed of canonical syllables.

Our particular predictions about effects of Risk did not, however, play out in the data. We observed no main effect of autism risk on CBRs. The finding adds further uncertainty to the already mixed evidence on canonical babbling emergence in autism and autism risk. The results support the argument that canonical babbling may be a robust developmental phenomenon and is more resistant to autism or autism risk than may have been previously assumed. Furthermore, and in contradiction to our initial expectation, we did not observe an overall tendency for CBRs to grow faster across Age in LR than HR infants. Instead, we found a tendency for CBRs of HR infants to grow faster in the first age interval (7.5 to 9.5 months) while CBRs of LR infants grew faster in the second (9.5 to 12 months). This pattern proved to be especially associated with segments where infants engaged in high VP, that is, when they were not vocalizing to other people, but vocalizing endogenously.

Overall main effects revealed, of course, the expected strong effect of Age on CBRs, a finding consistent with all prior longitudinal studies of canonical babbling. The present data do, on the other hand, provide new findings: we observed high CBRs in both Risk groups during segments with TT and during segments with high VP. The effect of TT was considerable, being equivalent to 2-3 months of growth in CBR, and the effect was even larger for high VP.[4]

Social Motivation in Early Infancy

The social motivation reasoning behind our predictions is based in the assumption that HR infants may present with a reduced experience of social reward compared to LR infants and thus demonstrate early differences (presumably reductions) in vocal performance during social interaction. The findings for CBRs during TT, however, suggest similar levels of social motivation in both groups, with both showing the tendency to produce higher CBRs during segments rated as having any TT compared to those rated as having no TT whatsoever. These findings suggest robustness of social motivations for infant vocalization. Our hypotheses were based on an expectation of anomalous development in HR infants, assuming social motivation for vocalization may break down in the presence of neurodevelopmental differences affecting social cognition. The results suggest a stronger mechanism where human infant vocal tendencies

[4] In order to determine CBR differences across all variables, we chose to analyze four statistical models. We were suspicious of results based on any single model after an initial full interaction model including TT and VP failed to converge, whereas the separate TT and VP models included enough Power for statistical significance to be evaluated. We therefore, proceeded with a more differentiated plan. In the first full model with TT, the only significant main effect we found was for one of the two age intervals, but surprisingly there was no significant main effect for TT. The lack of significance for TT may have been the result of high variance introduced by the Risk and Age factors and their interactions, with consequent reduction in the power to assess TT as a main effect. We are also uncertain of the possible role of the fact that 80% of segments were rated as having no TT whatsoever, and there was also a greatly unbalanced number of segments rated as 2-5 within the Any TT category—more than 80% were rated 2. The full model with VP, however, resulted in several significant effects including a highly significant main effect for VP. We are suspicious that the unexpected three-way interactions observed in the full models will not replicate. Consequently, we chose to conduct additional simplified analyses to provide perspective. The Age vs Risk model as well as the main effects model can be thought of as post hoc analyses in the service of the goal of descriptive perspective.

may have been selected to withstand the neurodevelopmental differences associated with autism risk.

There can be no doubt that humans are highly social beings. Clearly, early hominins' relatively large living groups necessitated a high level of social bonding, which created a need for an efficient communication method, and resulted in positive selection pressures on the evolution of language (Dunbar, 1993, 1996, 2004). Chevallier (2012) noted that "social motivation constitutes an evolutionary adaptation geared to enhance the individual's fitness in collaborative environments" (p. 2). Thus, it is reasonable to assume that precursors to language such as canonical babbling must be robust during development. Although often delayed in developmental disorders, including autism (Chericoni et al., 2016; Iverson & Wozniak, 2007; Patten et al., 2014), canonical babbling is well-established as a robust stage of development, known to emerge eventually even when infants cannot hear sounds produced in their environment, as in the case of deafness or severe hearing impairment (Eilers & Oller, 1994; Oller & Eilers, 1988). Our results indicated no overall difference between CBRs of HR and LR infants—only the patterns of growth in CBR appeared to differ—suggesting the quality of prelinguistic vocal forms (i.e., CBR) produced during early face-to-face interactions may be robust with respect to these evolutionary pressures.

One important consideration and potential limitation in this evaluation of social motivations in early infancy relates to the measures we used to assess the sociality of vocalizations. To measure infant TT, coders were asked to estimate on a Likert scale how often infants engaged in TT for each segment. This subjective measure, obtained immediately after coding for CBR for each segment, can be portrayed as a blunt instrument, subject to only fair inter-observer agreement, but it is founded in the notion that human judgments are the gold

standard for any such measure, and our method of obtaining the judgments was convenient and workable. A perhaps more reliable measure would require labeling the social or exploratory function of each utterance individually with repeat-observation (and especially with both audio and video), a measure that requires at least tenfold more time to obtain (see Long et al. (2020) for an analysis using this method). Future studies using this more expensive measure of the role of TT in infant vocalization are planned. An additional consideration includes examining infant-directed speech using similar methods employed in this study, as briefly discussed in Appendix A.

TT occurred, according to the coders, in only about 20% of the segments, a pattern that applied roughly equally to both Risk groups. This low rate of TT surprised us, given that so much of the literature on early language development focuses on protoconversation and its presumable importance in development. The low rate of TT may also have imposed a power limitation on the statistical analyses of the effects of TT and its interactions with the other variables in the present work.

Endogenous Motivation and Canonical Babbling

The VP measure was also based on a Likert scale, where coders were asked to judge each segment on how much of the time the infant had engaged in independent, not socially-directed vocalization (presumably endogenously motivated). Unlike TT, VP was found by the coders to be present in the vast majority of segments, and again this was true of both Risk groups—the plurality of segments having been rated 5 (VP present in close to the whole segment) by the coders for both Risk groups. Our surprise at low rates of TT in the all-day recordings is matched by our surprise at the near omnipresence of VP.

Again, however, the instrument measuring VP can be portrayed as blunt, having been obtained as a quick judgment from coders right after having completed listening to each segment and lacking high inter-coder agreement. The subjectivity of the judgments can be viewed in the same way as the TT judgments—human coding must be the gold standard in spite of its limitations. However, as with TT, more time-consuming judgments with audio and video and with repeat-observation coding are desirable.

Our hypotheses regarding VP were also based in part on the social motivation theory. We expected HR infants to show vocal behaviors similar to motoric behaviors that are characteristic of autism, such as frequent isolated play, stereotypic repetition of motoric behaviors, and preference for physical properties of objects (and thus acoustic-perceptual properties of sounds). Therefore, we anticipated HR infants would show a tendency to produce more canonical syllables than LR infants during high VP.

Overall, both the LR and HR groups produced more canonical syllables during high VP compared to low VP, but perhaps the most interesting outcome was the three-way interaction in the full VP model. The interaction suggests different rates of growth in CBR for HR and LR infants during the first and second age intervals (HR infants progressing faster in the first interval, LR infants faster in the second), but only for high VP segments. Low VP segments showed no such differentiation of Risk groups.

The social motivation theory posits that reduced early social reward processing affects later social cognitive functioning; however, Bottini (2018) described alternative hypotheses that have also been proposed to describe differences observed in autism, including general reward processing deficits in both social *and* non-social domains (Dichter et al., 2012; Kohls et al., 2013), and greater reward processing for non-social stimuli (Benning et al., 2016; Kohls et al.,

2014; Sasson et al., 2012). Our findings hint at the possibility that whatever the social motivation or reward systems are, they may function differentially at different points in time for infants at risk for autism and for infants not at risk. One might propose that HR infants may experience greater intrinsic reward when producing canonical syllables during bouts of vocal play (i.e., as non-social stimuli) compared to LR infants; yet this greater reward applied from 7.5 to 9.5 months, while dropping substantially from 9.5 to 12 months.

As previously mentioned, one of the primary diagnostic characteristics of autism is the presence of restricted interests and repetitive behaviors (RRBs), including repetitive movements with objects, repeated body movements, ritualistic behavior, restricted interests, and sensory sensitivities (American Psychiatric Association, 2013). RRBs are present in typically developing infants (Arnott et al., 2010), but occur more frequently in infants with autism than in neurotypical controls as young as 6 months of age (Richler et al., 2007; Rogers, 2009). High rates of canonical syllables observed during bouts of VP in these HR infants may represent manifestations of vocal stereotypies, similar to those seen in autism. It is thought that autistic infants may prefer playing with the sensorimotor characteristics of a syllable through repetition, while their neurotypical counterparts tend to play with varying aspects pertaining to individual syllables, modifying duration, placement, and various articulatory patterns from utterance to utterance. Thus, attending to the repetitive physical and acoustic properties of sounds during bouts of VP may be more intrinsically rewarding to infants with autism compared to typical development, who may be vocally exploring phonetic nuances. This idea is supported by the speech attunement framework (Shriberg et al., 2011), which proposes that autistic children process acoustic-perceptual characteristics more easily than semantic-linguistic information (Heaton et al., 2008; Järvinen-Pasley et al., 2008; Mottron et al., 2006).

Yet, the higher CBRs in HR infants compared to LR infants during high VP applied only for the first age interval, with an opposite pattern occurring thereafter (LR infants showing greater growth of CBR). Thus, if HR infants' increased CBRs in the first age interval are the result of autism-like repetition and stereotypy, there must be some other force at stake in the second age interval. Perhaps the robust tendency for canonical babbling to develop—based on the critical requirement for command of canonical syllables—drives all infants to reach a minimal level of canonical babbling control by the time word learning begins to take off at the end of the first year. Delays in the emergence rate of advanced vocal forms in infants at risk may become more evident at later ages as greater social and linguistic demands are placed on children who will show effects of autism. Such later delays may be foreshadowed in our finding of a plateauing of CBRs in HR infants by 12 months.

A potential limitation of this study is that we only evaluated the production of canonical babbling as a measure of advanced vocal forms without specifying the phonetic content of either canonical or non-canonical syllables. Infants are known to produce a wide range of vocal sounds throughout the first year. Given previous findings that RRBs are observed in infants as young as 6 months, infants with or at risk for autism may also produce non-canonical sounds as vocal or auditory self-stimulatory behaviors, sounds such as raspberries or simple vowel sounds. Once canonical babbling begins, phonetic characterization may be useful in supply details potentially relevant to prediction of language outcomes. A more in-depth evaluation of the production of prelinguistic sounds, both canonical and non-canonical, is necessary to better understand the emergence of vocal RRBs in autism.

On the Criterion for Canonical Babbling Onset

Canonical babbling onset has often been often suggested in the vocal developmental literature as requiring a .15 CBR based on a coded sample (i.e., 15% of syllables in a sample are canonical) (Lewedag, 1995; Nathani et al., 2006; Oller, 2000). However, this level appears to be untenable for the kinds of recordings and methods reported on in the present work. Our findings reveal that even the LR infants (who were presumably all typically developing) did not reach this criterion level until 12 months of age, despite expected mastery by 7-10 months. Previous research supporting the .15 criterion has primarily been completed in laboratory conditions (Molemans et al., 2012; Oller et al., 1994), settings in which parents have been expected to intentionally (or unintentionally) elicit and induce infant vocalization. Our findings and those discussed by Oller et al. (2020) support the notion that the average CBR in all-day recordings is lower than in laboratory settings.

Conclusions

The findings observed in the present study offer perspective on the ability to detect developmental differences in infant vocal turn taking and independent vocal production as potential indicators of autism. We observed a similar emergence of canonical babbling in infants at low and high risk for autism, and higher rates of canonical babbling overall during segments rated as having any turn taking and high vocal play. Our findings offer support for a potentially robust social motivation in infancy to produce higher rates of canonical syllables during interaction, even in the presence of possible social communication deficits. Differences observed between groups did occur when comparing low and high levels of independent vocal play. Evolutionary pressures may play a role in high-risk infants' increased rate of canonical syllables during vocal play early in the canonical babbling stage as a result of the need to signal fitness

prior to vocal delays at later ages. These differences may also support an age-varying heightened intrinsic reward mechanism for producing and attending to acoustic-perceptual characteristics of vocal sounds potentially linked to genes associated with autism.

5. Conclusion

The first study discussed in Chapter 2 examined the reliability of listener judgments of infant vocal imitation (Long et al., 2019). High intra- and inter-rater correlations were found among listeners on judgments of infant vocal imitativeness. Imitation was also observed to occur rarely, making up less than 5% of the total protophones. These confirmatory findings highlight the salience of vocal imitation, although infrequent in occurrence, supporting the claim that it may serve as a reliable fitness signal of infant communicative abilities.

Chapter 3 evaluated how often infants produce social and endogenous vocalizations across engaged (parent and infant interacting) and independent (parent interacting with another adult with baby in the room) laboratory circumstances (Long et al., 2020). Surprisingly, approximately 75% of all infant protophones across the second half year of life were produced endogenously with 67% in the engaged circumstance and 82% in the independent circumstance. These findings suggest that high rates of endogenously produced sounds may be an important indicator of fitness and support the claim that positive selection pressures may have been placed on the production of endogenous protophones as indicators of developmental well-being. Specifically, during vocal play, infants are learning the range of capabilities of their vocal system through sensorimotor exploration which parents can use to gauge developmental level. Furthermore, endogenously produced sounds provide the raw vocal material with which parents can engage during face-to-face interaction.

Chapter 4 (Long et al., in submission) evaluated rates of canonical syllables across various levels of infant turn taking and vocal play in infants at low and high risk for autism to study social and endogenous motivations involved in infant vocal production throughout the first year of life. The results showed no differences in rates of canonical syllables between risk groups during “no” and “any” turn taking, highlighting a potentially robust social motivation mechanism

in the face of social-communication impairments. There were also no significant differences observed in the rate of canonical syllables over time between risk groups, adding to the already mixed evidence on vocal developmental differences in autism (Lang et al., 2019). Finally, there was a significant difference between risk groups in the rate of canonical syllables produced during high vocal play. Specifically, high-risk infants produced more canonical syllables during high vocal play compared to the low-risk infants at the middle age studied. These findings suggest potential early predictors for vocal differences and potential communication impairments in autism and offer perspective on the fitness signaling hypothesis. Higher rates of canonical syllables in high-risk infants may be indicative of an intrinsic mechanism driving signaling wellness and thus, investment from caregivers. These infants may also experience greater intrinsic reward playing with the acoustic and articulatory properties of canonical syllables compared to low-risk infants, who may instead be attending to, and learning, the various capabilities of sound production in support of learning language to communicate (Heaton et al., 2008; Järvinen-Pasley et al., 2008; Mottron et al., 2006).

This dissertation evaluated the role of social and endogenous protophones as vocal fitness signals in typical and atypical human development and in the evolution of language. Specifically examining the salience of the vocal signal, overall proportions of social and endogenous vocalizations, and the role of advanced vocal forms and motivations in signaling wellness. Results from these studies elucidate the ways in which social and endogenous vocalizations can reveal natural selection pressures on fitness signaling in the human infant.

References

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2017). Multiple coordination patterns in infant and adult vocalization. *Infancy*, *22*(4), 514–539. https://doi.org/10.1111/infa.12165

Ainsworth, M. D. (1969). Object relations, dependency, and attachment: A theoretical review of the infant-mother relationship. *Child Development*, 969–1025. https://doi.org/10.1111/j.1467-8624.1969.tb04561.x

Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science*, *21*(5), 1–11. https://doi.org/10.1111/desc.12641

American Psychiatric Association. (2013). Autism Spectrum Disorder. In *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*.

Arazi, A., Censor, N., & Dinstein, I. (2017). Neural variability quenching predicts individual perceptual abilities. *Journal of Neuroscience*, *37*(1), 97–109. https://doi.org/10.1523/JNEUROSCI.1671-16.2016

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

Arnott, B., McConachie, H., Meins, E., Fernyhough, C., Couteur, A. Le, Turner, M., Parkinson, K., Vittorini, L., & Leekam, S. (2010). The frequency of restricted and repetitive behaviors in a community sample of 15-month-old infants. *Journal of Developmental and Behavioral Pediatrics*, *31*(3), 223–229. https://doi.org/10.1097/DBP.0b013e3181d5a2ad

Atlas, J. A. (1990). Play in assessment and intervention in the childhood psychoses. *Child Psychiatry and Human Development*, *21*(2), 199–133.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198245537.001.0001

Baranek, G. T., Watson, L. R., Boyd, B. A., Poe, M. D., David, F. J., & McGuire, L. (2013). Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Development and Psychopathology*. https://doi.org/10.1017/S0954579412001071

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, *19*(2), 159–170. https://doi.org/10.1016/s0163-6383(96)90015-6

Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113. https://doi.org/10.1111/j.1749-6632.1975.tb41575.x

Bekoff, M., & Byers, J. (1998). *Animal play: Evolutionary, comparative, and ecological perspectives*. Cambridge University.

Belardi, K., Watson, L. R., Faldowski, R. A., Hazlett, H., Crais, E., Baranek, G. T., McComish, C., Patten, E., & Oller, D. K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with Fragile X syndrome at 9 – 12 months of mge. *Journal of Autism and Developmental Disorders*, *47*(4), 1193–1206. https://doi.org/10.1177/0885066614530659.The

Benning, S. D., Kovac, M., Campbell, A., Miller, S., Hanna, E. K., Damiano, C. R., Sabatino-DiCriscio, A., Turner-Brown, L., Sasson, N. J., Aaron, R. V., Kinard, J., & Dichter, G. S. (2016). Late positive potential ERP Responses to social and nonsocial stimuli in youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(9), 3068–3077. https://doi.org/10.1007/s10803-016-2845-y

Berk, L. E. (1994). Vygotsky's theory: The importance of make-believe play. *Young Children*, *50*(1), 30–39.

Bertossa, R. C. (2011). Theme issue: Evolutionary developmental biology (evo-devo) and behaviour: Papers of a Theme issue compiled and edited by Rinaldo C. Bertossa. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 2055–2180. https://doi.org/10.1098/rstb.2011.0035

Bloom, K., & Esposito, A. (1975). Social conditioning and its proper control procedures. *Journal of Experimental Child Psychology*, *19*(2), 209–222. https://doi.org/10.1016/0022-0965(75)90085-5

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, *14*(2), 211–227. https://doi.org/10.1017/S0305000900012897

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, *6*(3), 380–420. https://doi.org/10.1016/0010-0285(74)90018-8

Bodnarchuk, J. L., & Eaton, W. O. (2004). Can parent reports be trusted? Validity of daily checklists of gross motor milestone attainment. *Applied Developmental Psychology*, *25*, 481–490. https://doi.org/10.1016/j.appdev.2004.06.005

Bornstein, M. H. (2000). Infant into conversant: Language and nonlanguage processes in developing early communication. In N. Budwig, I. Č. Užgiris, & J. V. Wertsch (Eds.), *Communication: An Arena of Development* (pp. 109–129). Greenwood Publishing Group.

Bottini, S. (2018). Social reward processing in individuals with autism spectrum disorder: A systematic review of the social motivation hypothesis. *Research in Autism Spectrum Disorders*, *45*, 9–26. https://doi.org/10.1016/j.rasd.2017.10.001

Bourvis, N., Singer, M., Saint Georges, C., Bodeau, N., Chetouani, M., Cohen, D., & Feldman, R. (2018). Pre-linguistic infants employ complex communicative loops to engage mothers

in social exchanges and repair interaction ruptures. *Royal Society Open Science*, *5*. https://doi.org/10.1098/rsos.170274

Bowlby, J. (1969). *Attachment and loss* (Vol. 1). Random House.

Bråten, S. (1988). Dialogic mind: The infant and the adult in protoconversation. In M. E. Carvallo (Ed.), *Nature, Cognition and System I. Theory and Decision Library (Series D: System Theory, Knowledge Engineering and Problem Solving), vol 2* (pp. 187–205). Springer. https://doi.org/10.1007/978-94-009-2991-3_9

Breland, H. M. (1974). Birth order, family configuration, and verbal achievement. *Child Development*, *45*(4), 1011. https://doi.org/10.2307/1128089

Bruner, J. S. (1974). From communication to language- A psychological perspective. *Cognition*, *3*(3), 255–287. https://doi.org/10.1016/0010-0277(74)90012-2

Buder, E. H., Chorna, L. B., Oller, D. K., & Robinson, R. B. (2008). Vibratory regime classification of infant phonation. *Journal of Voice : Official Journal of the Voice Foundation*, *22*(5), 553–564. https://doi.org/10.1016/j.jvoice.2006.12.009

Buder, E. H., Warlaumont, A. S., Oller, D. K., & Chorna, L. B. (2010). Dynamic indicators of mother-infant prosodic and illocutionary coordination. *Speech Prosody 2010-Fifth International Conference*, 6–9.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Caligiore, D., Ferrauto, T., Parisi, D., Accornero, N., Capozza, M., & Baldassarre, G. (2008). Using motor babbling and Hebb rules for modeling the development of reaching with obstacles and grasping. *International Conference on Cognitive Systems*, E1-8.

Camarata, S., & Yoder, P. (2002). Language transactions during development and intervention: Theoretical implications for developmental neuroscience. In *International Journal of Developmental Neuroscience* (Vol. 20, Issues 3–5, pp. 459–465). Elsevier Ltd. https://doi.org/10.1016/S0736-5748(02)00044-8

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- to 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330. https://doi.org/10.1016/s0163-6383(98)90009-1

Carroll, S. B. (2005). *Endless forms most beautiful: The new science of evo-devo and the making of the animal kingdom*. W. W. Norton & Co. https://doi.org/10.1086/503946

Chericoni, N., de Brito Wanderley, D., Costanzo, V., Diniz-Gonçalves, A., Gille, M. L., Parlato, E., Cohen, D., Apicella, F., Calderoni, S., & Muratori, F. (2016). Pre-linguistic vocal trajectories at 6-18 months of age as early markers of autism. *Frontiers in Psychology*, *7*(OCT), 1595. https://doi.org/10.3389/fpsyg.2016.01595

Chevallier, C., Huguet, P., Happé, F., George, N., & Conty, L. (2013). Salient social cues are prioritized in autism spectrum disorders despite overall decrease in social attention. *Journal of Autism and Developmental Disorders*, *43*(7), 1642–1651. https://doi.org/10.1007/s10803-012-1710-x

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Clark, R. (1977). What's the use of imitation? *Journal of Child Language*, *4*(3), 341–358. https://doi.org/10.1017/S0305000900001732

Collie, R., & Hayne, H. (1999). Deferred imitation by 6- and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology*, *35*(2), 83–90. https://doi.org/doi.org/10.1002/(SICI)1098-2302(199909)35:2%3C83::AID-DEV1%3E3.0.CO;2-S

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology*, *58*(1), 175–199. https://doi.org/10.1146/annurev.psych.58.110405.085551

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*, 759–773. https://doi.org/10.1111/cdev.12974

Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., & Jaffe, J. (2002). The cross-modal coordination of interpersonal timing: Six-week-olds infants' gaze with adults' vocal behavior. *Journal of Psycholinguistic Research*, *31*(1), 1–23. https://doi.org/10.1023/A:1014301303616

Darwin, C. (1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. John Murray. https://doi.org/10.5962/bhl.title.68064

Davis, K. L., & Panksepp, J. (2018). *The emotional foundations of personality: A neurobiological and evolutionary approach*. W.W. Norton & Company.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.40.2.271

Deacon, T. W. (1997). *The symbolic species*. W. W. Norton & Co. https://doi.org/10.1093/ww/9780199540884.013.u213400

Delgado, C. E. F., Messinger, D. S., & Yale, M. E. (2002). Infant responses to direction of parental gaze: A comparison of two still-face conditions. *Infant Behavior and Development*, *25*(3), 311–318. https://doi.org/10.1016/S0163-6383(02)00096-6

Delgado, R. E., Buder, E. H., & Oller, D. K. (2010). *AACT (Action Analysis Coding and Training)*. Intelligent Hearing Systems.

Dichter, G. S., Felder, J. N., Green, S. R., Rittenberg, A. M., Sasson, N. J., & Bodfish, J. W. (2012). Reward circuitry function in autism spectrum disorders. *Social Cognitive and Affective Neuroscience*, *7*(2), 160–172. https://doi.org/10.1093/scan/nsq095

Dissanayake, E. (1992). *Homo aestheticus: Where art comes from and why*. University of Washington Press. https://doi.org/10.1525/aa.1993.95.1.02a00370

Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant and Child Development*, *25*(3), 240–255. https://doi.org/10.1002/icd.1976

Dunbar, R. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*, 681–735. https://doi.org/10.1017/s0140525x00032325

Dunbar, R. M. (1996). *Gossiping, grooming and the evolution of language*. Harvard University Press.

Dunbar, R. M. (2004). Language, music, and laughter in evolutionary perspective. In D. K. Oller & U. Griebel (Eds.), *The Evolution of Communication Systems: A Comparative Approach* (pp. 257–274). MIT Press. https://doi.org/10.7551/mitpress/2879.003.0021

Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, *124*, 199–203.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*, 998–1011. https://doi.org/10.1017/S0305000919000291

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*, 491–541. https://doi.org/10.1017/s0140525x04000111

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–266. https://doi.org/10.1016/s1364-6613(00)01494-7

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C.-C., Lambert, B., & Oller, D. K. (2013). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, *10*(3), 279–296. https://doi.org/10.1080/15475441.2013.849176

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/s1364-6613(98)01262-5

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In D. K. Oller & U. Griebel (Eds.), *The Evolution of Communication Systems: A Comparative Approach*

(pp. 237–256). MIT Press. https://doi.org/10.7551/mitpress/2879.001.0001

Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology*, *67*(9), 10.1007/s00265-013-1491-z. https://doi.org/10.1007/s00265-013-1491-z

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Gleason, J. R., & Ely, R. (2002). Gender differences in language development. In A. V. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, Society, and Behavior: The Development of Sex Differences in Cognition* (Vol. 21). Greenwood Publishing Group.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the USA*, *100*(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

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 caregivers. *Child Development*, *80*(3), 636–644. https://doi.org/10.1111/j.1467-8624.2009.01287.x

Golinkoff, Roberta M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Golinkoff, Roberta Michnick, Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) Talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gottlieb, G. (2002). Developmental-behavioral initiation of evolutionary change. *Psychological Review*, *109*(2), 211. https://doi.org/10.1037/0033-295X.109.2.211

Gratier, M., & Devouche, E. (2011). Imitation and Repetition of Prosodic Contour in Vocal Interaction at 3 Months. *Developmental Psychology*, *47*(1), 67–76. https://doi.org/10.1037/a0020722

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*(September), 1–10. https://doi.org/10.3389/fpsyg.2015.01167

Gray, K. L. H., Haffey, A., Mihaylova, H. L., & Chakrabarti, B. (2018). Lack of privileged

access to awareness for rewarding social scenes in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(10), 3311–3318. https://doi.org/10.1007/s10803-018-3595-9

Griebel, U., & Oller, D. K. (2008). Evolutionary forces favoring contextual flexibility. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility: Complexity, Creativity and Adaptability in Human and Animal Communication* (pp. 9–40). MIT Press. https://doi.org/10.7551/mitpress/9780262151214.003.0002

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516. https://doi.org/10.1177/0165025406071914

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19*(4), 385–408. https://doi.org/10.1111/infa.12054

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19*(4), 385–408.

Grossberg, S., & Vladusich, T. (2010). How do children learn to follow gaze, share joint attention, imitate their teachers, and use tools during social interactions? *Neural Networks*, *23*(8–9), 940–965. https://doi.org/10.1016/j.neunet.2010.07.011

Gruss, L. T., & Schmitt, D. (2015). The evolution of the human pelvis: Changing adaptations to bipedalism, obstetrics and thermoregulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1663), 20140063. https://doi.org/10.1098/rstb.2014.0063

Harold, M. P., & Barlow, S. M. (2013). Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. *Infant Behavior and Development*, *36*(1), 84–93. https://doi.org/10.1016/j.infbeh.2012.10.001

Hastings, R. P., Kovshoff, H., Ward, N. J., Espinosa, F. degli, Brown, T., & Remington, B. (2005). Systems analysis of stress and positive perceptions in mothers and fathers of pre-school children with autism. *Journal of Autism and Developmental Disorders*, *35*(5), 635–644. https://doi.org/10.1007/s10803-005-0007-8

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(6), 771–782. https://doi.org/10.1080/02643290802336277

Heimann, M., Edorsson, A., Sundqvist, A., & Koch, F. (2017). Thirteen- to sixteen-months old infants are able to imitate a novel act from memory in both unfamiliar and familiar settings

but do not show evidence of rational inferential processes. *Frontiers in Psychology*, *8*(December), 2186. https://doi.org/10.3389/fpsyg.2017.02186

Heimann, M., Nelson, K. E., & Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: Methodological aspects and evidence of early individual differences. *Scandinavian Journal of Psychology*, *30*(2), 90–101. https://doi.org/10.1111/j.1467-9450.1989.tb01072.x

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*(4), 603–629. https://doi.org/10.1017/s0142716400010389

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hsu, H. C., & Fogel, A. (2001). Infant vocal development in a dynamic mother-infant communication system. *Infancy*, *2*(1), 87–109. https://doi.org/10.1207/S15327078IN0201_6

Hsu, H. C., & Fogel, A. (2003). Social regulatory effects of infant nondistress vocalization on maternal behavior. *Developmental Psychology*, *39*(6), 976. https://doi.org/10.1037/0012-1649.39.6.976

Hurford, J. R. (2011). *The origins of grammar*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0040

Hurley, S., & Chater, N. (2005a). *Perspectives on imitation: From neuroscience to social science - Volume 1: Mechanisms of imitation and imitation in animals.* The MIT Press.

Hurley, S., & Chater, N. (2005b). *Perspectives on imitation: From neuroscience to social science - Volume 2: Imitation, human development, and culture.* The MIT Press.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., Lyons, T., & Al, E. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science*, *22*(6). https://doi.org/10.1111/desc.12825

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders*, *37*(1), 158–170. https://doi.org/10.1007/s10803-006-0339-z

Iyer, S. N., Denson, H., Lazar, N., & Oller, D. K. (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). *Clinical Linguistics and Phonetics*, *30*(6), 470–788. https://doi.org/10.3109/02699206.2016.1147082

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*(2). https://doi.org/10.2307/3181589

Järvinen-Pasley, A., Wallace, G. L., Ramus, F., Happé, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science*, *11*(1), 109–121. https://doi.org/10.1111/j.1467-7687.2007.00644.x

Jhang, Y., Franklin, B., Ramsdell-Hudock, H. L., & Oller, D. K. (2017). Differing roles of the face and voice in early human communication: Roots of language in multimodal expression. *Frontiers in Communication*, *2*(10), 10. https://doi.org/10.3389/fcomm.2017.00010

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology*, *8*, 300. https://doi.org/10.3389/fpsyg.2017.00300

Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science*, *18*(7), 593–599. https://doi.org/10.1111/j.1467-9280.2007.01945.x

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions: Biological Sciences*, *364*, 2325–2335.

Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2013). Deficit, difference, or both? Autism and neurodiversity. *Developmental Psychology*, *49*(1), 59–71. https://doi.org/10.1037/a0028353

Kellerman, A. M., Schwichtenberg, A. J., Tonnsen, B. L., Posada, G., & Lane, S. P. (2019). Dyadic interactions in children exhibiting the broader autism phenotype: Is the broader autism phenotype distinguishable from typical development? *Autism Research*, *12*(3), 469–481. https://doi.org/10.1002/aur.2062

Kerig, P. K., Cowan, P. A., & Cowan, C. P. (1993). Marital quality and gender differences in parent-child interaction. *Developmental Psychology*, *29*(6), 931–939. https://doi.org/10.1037/0012-1649.29.6.931

Keven, N., & Akins, K. A. (2016). Neonatal imitation in context: Sensorimotor development in the perinatal period. *Behavioral and Brain Sciences*, *40*(2017), 1–107. https://doi.org/10.1017/s0140525x16000911

Kirschner, M., & Gerhart, J. (2006). *The plausibility of life: Resolving Darwin's dilemma*. Yale University.

Klein, P. J., & Meltzoff, A. N. (1999). Long-term memory, forgetting, and deferred imitation in 12-month-old infants. *Developmental Science*, *2*(1), 102–113. https://doi.org/10.1111/1467-7687.00060

Kohls, G., Schulte-Rüther, M., Nehrkorn, B., Müller, K., Fink, G. R., Kamp-Becker, I., Herpertz-Dahlmann, B., Schultz, R. T., & Konrad, K. (2013). Reward system dysfunction in autism

spectrum disorders. *Social Cognitive and Affective Neuroscience*, *8*(5), 565–572. https://doi.org/10.1093/scan/nss033

Kohls, G., Thönessen, H., Bartley, G. K., Grossheinrich, N., Fink, G. R., Herpertz-Dahlmann, B., & Konrad, K. (2014). Differentiating neural reward responsiveness in autism versus ADHD. *Developmental Cognitive Neuroscience*, *10*, 104–116. https://doi.org/10.1016/j.dcn.2014.08.003

Kojima, S. (2003). A search for the origins of human speech: Auditory and vocal functions of the chimpanzee. *ISBS*. https://doi.org/10.1159/isbn.978-3-318-01930-8

Koopmans-van Beinum, F. J., & van der Stelt, J. M. (1986). Early stages in the development of speech movements. In *Precursors of Early Speech* (pp. 37–50). Palgrave Macmillan UK.

Kugiumutzakis, G. (1999). Genesis and development of early infant mimesis to facial and vocal models. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 36–59). Cambridge University Press.

Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, *10*(1), 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4 Pt 1), 2425–2438. https://doi.org/10.1121/1.417951

Lafreniere, P. (2011). Evolutionary functions of social play: Life histories, sex differences, and emotion regulation. *American Journal of Play*, *3*(4), 464–488.

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports*, *6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

Latta, R. G. (2010). Natural selection, variation, adaptation, and evolution: A primer of interrelated concepts. *International Journal of Plant Sciences*, *171*(9), 930–944. https://doi.org/10.1086/656220

LeBarton, E. S., & Iverson, J. M. (2016). Associations between gross motor and communicative development in at-risk infants. *Infant Behavior and Development*, *44*, 59–67. https://doi.org/10.1016/j.infbeh.2016.05.003

Lee, C.-C., Jhang, Y., Relyea, G., Chen, L.-M., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development*, *50*, 140–153. https://doi.org/10.1016/j.infbeh.2017.12.002

Leonard, L. B., Schwartz, R. G., Folger, M. K., Newhoff, M., & Wilcox, M. J. (1979). Children's imitations of lexical items. *Child Development*, *50*(1), 19–27. https://doi.org/10.1111/j.1467-8624.1979.tb02974.x

Levin, K. (1999). Babbling in infants with cerebral palsy. *Clinical Linguistics & Phonetics*, *13*(4), 249–267. https://doi.org/10.1080/026992099299077

Lewedag, V. L. (1995). *Patterns of onset of canonical babbling among typically developing infants*. University of Miami.

Lewis, M. M. (1936). *Infant Speech*. Harcourt Brace.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Locke, J. L. (2006). Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. *Human Nature*, *17*, 155–168. https://doi.org/10.1007/s12110-006-1015-x

Locke, J. L. (2008). Lipsmacking and babbling: Syllables, sociality, and survival. In B. L. Davis & K. Zajdo (Eds.), *The Syllable in Speech Production*. Erlbaum. https://doi.org/10.4324/9780203837894

Locke, J. L. (2009). Evolutionary developmental linguistics: Naturalization of the faculty of language. *Language Sciences*, *31*, 33–59. https://doi.org/10.1016/j.langsci.2007.09.008

Locke, J. L. (2017). Emancipation of the voice: Vocal complexity as a fitness indicator. *Psychonomic Bulletin & Review*, *24*(1), 232–237.

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*, 259–325.

Lohmander, A., Holm, K., Eriksson, S., & Lieberman, M. (2017). Observation method identifies that a lack of canonical babbling can indicate future speech and language problems. *Acta Paediatrica, International Journal of Paediatrics*, *106*(6), 935–943. https://doi.org/10.1111/apa.13816

Long, H. L., Bowman, D. D., Yoo, H., Burkhardt-Reed, M. M., Bene, E. R., & Oller, D. K. (2020). Social and endogenous infant vocalizations. *PLoS ONE*, *15*(8), e0224956. https://doi.org/10.1371/journal.pone.0224956

Long, H. L., Oller, D. K., & Bowman, D. D. (2019). Reliability of listener judgments of infant vocal imitation. *Frontiers in Psychology*, *10*, 1340. https://doi.org/10.3389/fpsyg.2019.01340

Long, H. L., Oller, D. K., Ramsdell-Hudock, H. L., & Bene, E. (2016). Imitative and vocally adaptive behaviors in infants through twelve months. *Annual Meeting of the American Speech-Language Hearing Association*.

Long, H. L., Ramsay, G., Bowman, D. D., Burkhardt-Reed, M. M., & Oller, D. K. (in submission). *Social and endogenous motivations in the emergence of canonical babbling:*

An autism risk study.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

Lynch, M. P., Oller, D. K., Steffens, M. L., Levine, S. L., Basinger, D. L., & Umbel, V. (1995). Onset of speech-like vocalizations in infants with Down syndrome. *American Journal of Mental Retardation*, *100*(1), 68–86.

Lyytinen, P., Poikkeus, A. M., Leiwo, M., Ahonen, T., & Lyytinen, H. (1996). Parents as informants of their child's vocal and early language development. *Early Child Development and Care*, *126*(1), 15–25. https://doi.org/10.1080/0300443961260102

Masataka, N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: Late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling. *Developmental Science*. https://doi.org/10.1111/1467-7687.00161

Masur, E. F. (2006). Vocal and action imitation by infants and toddlers during dyadic interactions. In S. J. Rogers & J. H. G. Williams (Eds.), *Imitation and the social mind: Autism and typical development.* (pp. 27–47). The Guilford Press.

Masur, E. F., & Eichorst, D. L. (2002). Infants' spontaneous imitation of novel versus familiar words: Relations to observational and maternal report measures of their lexicons. *Merrill-Palmer Quarterly*, *48*(4), 405–426. https://doi.org/10.1353/mpq.2002.0019

Meltzoff, A. N. (1988a). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development*, *59*(1), 217. https://doi.org/10.2307/1130404

Meltzoff, A. N. (1988b). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*(4), 470–476. https://doi.org/10.1037/0012-1649.24.4.470

Meltzoff, A. N. (2005). Imitation and other minds: The "Like Me" hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science* (Vol. 2, pp. 55–77). MIT Press.

Meltzoff, A. N. (2007). "Like me": a foundation for social cognition. *Developmental Science*, *10*(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75–78. https://doi.org/10.1126/science.198.4312.75

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*(3), 702. https://doi.org/10.2307/1130058

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, *25*(6), 954–962. https://doi.org/10.1037/0012-1649.25.6.954

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting*, *6*(3–4), 179–192. https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Meltzoff, A. N., & Moore, M. K. (2002). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, *25*(1), 39–61. https://doi.org/10.1016/s0163-6383(02)00090-5

Milenkovic, P. (2001). *TF32 [Computer software]*.

Moerk, E. L., & Moerk, C. (1979). Quotations, imitations, and generalizations. Factual and methodological analyses. *International Journal of Behavioral Development*, *2*(1), 43–72. https://doi.org/10.1177/016502547900200103

Molemans, I., Van den Verg, R., Van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language*, *39*(3), 523–552. https://doi.org/10.1017/S0305000911000171

Morris, D. (1967). *The naked ape*. Dell. https://doi.org/10.1002/bs.3830130309

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

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*, 1006. https://doi.org/10.3389/fpsyg.2013.01006

Moulin-Frier, C., & Oudeyer, P. Y. (2013). The role of intrinsic motivations in learning sensorimotor vocal mappings: A developmental robotics study. *INTERSPEECH, ISCA*.

Mowrer, O. H. (1960). *Learning theory and the symbolic processes.* John Wiley & Sons Inc. https://doi.org/10.1037/10772-000

Müller, G. B., & Newman, S. A. (2003). *Origination of organismal form: Beyond the gene in developmental and evolutionary biology*. MIT Press. https://doi.org/10.7551/mitpress/5182.001.0001

Mundy, P. (2017). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, 1–18. https://doi.org/10.1111/ejn.13720

Naber, F. B. A., Bakermans-Kranenburg, M. J., Van Ijzendoorn, M. H., Swinkels, S. H. N.,

Buitelaar, J. K., Dietz, C., Van Daalen, E., & Van Engeland, H. (2008). Play behavior and attachment in toddlers with autism. *Journal of Autism and Developmental Disorders*, *38*(5), 857–866. https://doi.org/10.1007/s10803-007-0454-5

Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics and Phonetics*, *20*(5), 351–369. https://doi.org/10.1080/02699200500211451

Newman, S. A. (2000). The role of genetic reductionism in biocolonialism. *Peace Review: A Journal of Social Justice*, *12*(4), 517–524. https://doi.org/10.1080/10402650020014592

Newman, S. A. (2012). Physico-genetic determinants in the evolution of development. *Science*, *338*, 217–219. https://doi.org/10.1126/science.1224311

Newman, S. A. (2016). Origination, variation, and conservation of animal body plan development. *Cell Biology and Molecular Medicine Reviews*, *2*, 130–162. https://doi.org/10.1002/3527600906.mcb.200400164

Nyman, A., & Lohmander, A. (2018). Babbling in children with neurodevelopmental disability and validity of a simplified way of measuring canonical babbling ratio. *Clinical Linguistics and Phonetics*, *32*(2), 114–127. https://doi.org/10.1080/02699206.2017.1320588

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Volume 1, Production* (pp. 93–1123). Academic Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L. B., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C.-C., Bowman, D. D., Long, H. L., Buder, E. H., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports*, *9*, 14734. https://doi.org/10.1038/s41598-019-51352-0

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441–449. https://doi.org/10.1111/j.1467-8624.1988.tb01479.x

Oller, D. K., Eilers, R. E., & Basinger, D. (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science*, *4*, 49–60. https://doi.org/10.1111/1467-7687.00148

Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, *103*(3), 249. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2

Oller, D. K., Eilers, R. E., Steffens, M. L., Urbano, R., & Lynch, M. P. (1994). Speech-Like Vocalizations in Infancy: An Evaluation of Potential Risk Factors. *Journal of Child Language*, *21*(1), 33–58. https://doi.org/10.1017/S0305000900008667

Oller, D. K., & Griebel, U. (2005). Contextual freedom in human infant vocalization and the evolution of language. In R. L. Burgess & K. MacDonald (Eds.), *Evolutionary Perspectives on Human Development* (pp. 135–166). SAGE Publications. https://doi.org/10.4135/9781452233574.n5

Oller, D. K., & Griebel, U. (2008). Contextual 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). MIT Press. https://doi.org/10.7551/mitpress/9780262151214.003.0007

Oller, D. K., Griebel, U., Bowman, D. D., Bene, E. R., Long, H. L., Yoo, H., & Ramsay, G. (2020). Infant boys are more vocal than infant girls. *Current Biology*, *30*, R417-429. https://doi.org/10.1016/j.cub.2020.03.049

Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in Psychology*, *10*, 729. https://doi.org/10.3389/fpsyg.2019.00729

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal development as a guide to modeling the evolution of language. *Topics in Cognitive Science*, *8*(2), 382–392. https://doi.org/10.1111/tops.12198.Vocal

Oller, D. K., Wieman, L. A., Doyle, J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*, 1–11.

Oostenbroek, J., Slaughter, V., Nielsen, M., & Suddendorf, T. (2013). Why the confusion around neonatal imitation? A review. *Journal of Reproductive and Infant Psychology*, *31*(4), 328–341. https://doi.org/10.1080/02646838.2013.832180

Ozonoff, S., Iosif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(3), 256-66.e1-2.

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*(3), 407–422. https://doi.org/10.1017/S0140525X00012759

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J. (2011). Toward a cross-species neuroscientific understanding of the affective mind: Do animals have emotional feelings? In *American Journal of Primatology* (Vol. 73, Issue 6, pp. 545–561). https://doi.org/10.1002/ajp.20929

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions*. W.V. Norton & Company. https://doi.org/10.5860/choice.50-3555

Panksepp, J., Siviy, S., & Normansell, L. (1984). The psychobiology of play: Theoretical and methodological perspectives. *Neuroscience and Biobehavioral Reviews*, *8*(4), 465–492. https://doi.org/10.1016/0149-7634(84)90005-8

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*(6), 137–158. https://doi.org/10.1177/014272378900900603

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, 1–16. https://doi.org/10.1007/s10803-014-2047-4

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(5), 588–598. https://doi.org/10.1111/j.1469-7610.2010.02332.x

Pawlby, S. J. (1977). Infant interaction. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 203–224). Academic Press.

Pedersen, F. A., Rubenstein, J. L., & Yarrow, L. J. (1979). Infant development in father-absent families. *Journal of Genetic Psychology*, *135*(1), 51–61. https://doi.org/10.1080/00221325.1979.10533416

Piaget, J. (1952a). *Play, dreams and imitation in childhood*. W. W. Norton & Co. https://doi.org/10.4324/9781315009698

Piaget, J. (1952b). The second stage: The first acquired adaptations and the primary circular reaction. In J. Piaget & M. Cook (Eds.), *The origins of intelligence in children* (pp. 47–143). W. W. Norton & Co. https://doi.org/10.1037/11494-003

Pipp, S., & Harmon, R. J. (1987). Attachment as regulation: A commentary. *Child Development*, *58*(3), 648–652. https://doi.org/10.2307/1130204

Pokorny, F. B., Schuller, B. W., Marschik, P. B., Brueckner, R., Nyström, P., Cummins, N., Bölte, S., Einspieler, C., & Falck-Ytter, T. (2017). Earlier identification of children with autism spectrum disorder: An automatic vocalisation-based approach. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2017-Augus*, 309–313. https://doi.org/10.21437/Interspeech.2017-1007

Ramer, A. L. H. (1976). The function of imitation in child language. *Journal of Speech and Hearing Research*, *19*(4), 700–717. https://doi.org/10.1044/jshr.1904.700

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, *14*(1), 92–105. https://doi.org/10.1111/j.1467-7687.2010.00961.x

Réger, Z. (1986). The functions of imitation in child language. *Applied Psycholinguistics*, *7*(4), 323–352. https://doi.org/10.1017/s0142716400007712

Richler, J., Bishop, S. L., Kleinke, J. R., & Lord, C. (2007). Restricted and repetitive behaviors in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(1), 73–85. https://doi.org/10.1007/s10803-006-0332-6

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Robson, S. L., van Schaik, C. P., & Hawkes, K. (2006). The derived features of human life history. In K. Hawkes & R. R. Paine (Eds.), *The Evolution of Human Life History* (pp. 1–16). School for Advanced Research Press.

Rochat, P., Querido, J. G. Q., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology*, *35*(4), 950–957.

Rodgon, M. M., & Kurdek, L. A. (1977). Vocal and gestural imitation in 8-, 14-, and 20-month-old children. *Journal of Genetic Psychology*, *131*(1), 115–123. https://doi.org/10.1080/00221325.1977.10533280

Rogers, S. J. (2009). What are infant siblings teaching us about Autism in infancy? *Autism Research*, *2*(3), 125–137. https://doi.org/10.1002/aur.81

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956–970. https://doi.org/10.1111/cdev.12166

Sakkalou, E., Ellis-Davies, K., Fowler, N. C., Hilbrink, E. E., & Gattis, M. (2013). Infants show stability of goal-directed imitation. *Journal of Experimental Child Psychology*, *114*(1), 1–9. https://doi.org/10.1016/j.jecp.2012.09.005

Sasson, N. J., Dichter, G. S., & Bodfish, J. W. (2012). Affective responses by adults with autism are reduced to social images but elevated to images related to circumscribed interests. *PLoS ONE*, *7*(8), e42457. https://doi.org/10.1371/journal.pone.0042457

Schore, A. N. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal*, *22*(1–2), 7–66.

https://doi.org/10.1002/1097-0355(200101/04)22:1<7::AID-IMHJ2>3.0.CO;2-N

Schreibman, L. (2005). *The science and fiction of autism*. Harvard University.

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D. J., & Gore, J. C. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. *Archives of General Psychiatry*, *57*, 331–340. https://doi.org/10.1001/archpsyc.57.4.331

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., & Bookheimer, S. Y. (2010). Reward Processing in Autism. *Autism Research*, *3*(2), 53–67. https://doi.org/10.1002/aur.122

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge University.

Sepeta, L., Tsuchiya, N., Davies, M. S., Sigman, M., Bookheimer, S. Y., & Dapretto, M. (2012). Abnormal social reward processing in autism as indexed by pupillary responses to happy faces. *Journal of Neurodevelopmental Disorders*, *4*(1), 1–9. https://doi.org/10.1186/1866-1955-4-17

Sheya, A., & Smith, L. B. (2013). Development through sensorimotor coordination. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction: Toward a new paradigm for cognitive science* (pp. 123–143). MIT Press. https://doi.org/10.7551/mitpress/9780262014601.003.0005

Shriberg, L. D., Paul, R., Black, L. M., & Van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(4), 405–426. https://doi.org/10.1007/s10803-010-1117-5

Siegal, M. (1987). Are sons and daughters treated more differently by fathers than by mothers? *Developmental Review*, *7*(3), 183–209. https://doi.org/10.1016/0273-2297(87)90012-8

Sigman, M., & Ungerer, J. A. (1984). Attachment behaviors in autistic children. *Journal of Autism and Developmental Disorders*, *14*(3), 231–244. https://doi.org/10.1007/BF02409576

Simpson, E. A., Paukner, A., Suomi, S. J., & Ferrari, P. F. (2015). Neonatal imitation and its sensorimotor mechanism. *New Frontiers in Mirror Neurons Research*, 296–314. https://doi.org/10.1093/acprof:oso/9780199686155.003.0016

Sinha, C. (2004). The evolution of language: From signals to symbols to system. In D. K. Oller & U. Griebel (Eds.), *The Evolution of Communication Systems: A Comparative Approach* (pp. 217–235). MIT Press.

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts, Inc.

Snow, C. E. (1989). Imitativeness: A trait or a skill? In G. E. Speidel & K. E. Nelson (Eds.), *The Many Faces of Imitation in Language Learning* (pp. 73–90). Springer.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds.), *Child Phonology* (Vol. 1, pp. 73–90). Academic Press.

Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Mental Health Journal, 2*(2), 118–128. https://doi.org/10.1002/1097-0355(198122)2:2<118::AID-IMHJ2280020208>3.0.CO;2-5

Su, P. L., Rogers, S. J., Estes, A. M., & Yoder, P. J. (2020). The role of early social motivation in explaining variability in functional language in toddlers with ASD. *Autism: The International Journal of Research & Practice*. https://doi.org/10.1177/1362361320953260

Sundqvist, A., Nordqvist, E., Koch, F.-S., & Heimann, M. (2016). Early declarative memory predicts productive language: A longitudinal study of deferred imitation and communication at 9 and 16months. *Journal of Experimental Child Psychology, 151*, 109–119. https://doi.org/10.1016/j.jecp.2016.01.015

Syal, S. (2011). *Socially motivated vocal learning.* [Doctoral dissertation, Cornell University]. https://doi.org/10.16194/j.cnki.31-1059/g4.2011.07.016

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). Cambridge University.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective Communication and Emotion in Early Ontogeny* (pp. 15–46). Cambridge University.

Tronick, E. Z., Als, H., Adamson, L. B., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1

Užgiris, I. Č. (2010). Imitation as activity: Its developmental aspects. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (Cambridge, pp. 186–206). Cambridge University Press.

Užgiris, I. Č., Benson, J. B., Kruper, J. C., & Vasek, M. E. (1989). Contextual influences on imitative interactions between mothers and infants. In *Action in Social Context* (pp. 103–127). Springer US. https://doi.org/10.1007/978-1-4757-9000-9_4

Vauclair, J., & Bard, K. A. (1983). Development of manipulations with objects in ape and human infants. *Journal of Human Evolution*, *12*(7), 631–645. https://doi.org/10.1016/S0047-2484(83)80003-7

Vygotsky, L. S. (1978). Interaction between learning and development. In *Mind in Society* (pp. 79–91). Harvard University Press.

Warlaumont, A. S., Oller, D. K., Buder, E. H., Dale, R., & Kozma, R. (2010). Data-driven

automated acoustic analysis of human infant vocalizations using neural network tools. *The Journal of the Acoustical Society of America*, *127*(4), 2563–2577. https://doi.org/10.1121/1.3327460

Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(32), 121–126.

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*, 1314–1324. https://doi.org/10.1177/0956797614531023

Washburn, S. L. (1960). Tools and human evolution. *Scientific American*, *203*(3), 62–75.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Weeks, S. J., & Hobson, R. P. (1987). The salience of facial expression for autistic children. *Journal of Child Psychology and Psychiatry*, *28*(1), 137–152. https://doi.org/10.1111/j.1469-7610.1987.tb00658.x

Wells, J. C. K., DeSilva, J. M., & Stock, J. T. (2012). The obstetric dilemma: An ancient game of Russian roulette, or a variable dilemma sensitive to ecology? *American Journal of Physical Anthropology*, *149*, 40–71. https://doi.org/10.1002/ajpa.22160

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders*, *30*(2), 157–162. https://doi.org/10.1023/A:1005463707029

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Wiley, A. S., & Allen, J. S. (2017). *Medical Anthropology: A Biocultural Approach* (3rd ed.). Oxford University Press.

Wiley, A. S., & Cullin, J. M. (2020). Biological normalcy. *Evolution, Medicine and Public Health*, *2020*(1), 1. https://doi.org/10.1093/emph/eoz035

Williams, E., Reddy, V., & Costall, A. (2001). Taking a closer look at functional play in children with autism. *Journal of Autism and Developmental Disorders*, *31*(1), 67–77. https://doi.org/10.1023/A:1005665714197

Winnicott, D. W. (1960). The theory of the parent-infant relationship. *The International Journal of Psychoanalysis*, *41*, 585–595.

Yoo, H., Bowman, D. D., & Oller, D. K. (2018). The origin of protoconversation: An examination of caregiver responses to cry and speech-like vocalizations. *Frontiers in Psychology*, *9*, 1510. https://doi.org/10.3389/fpsyg.2018.01510

Zajonc, R. B., & Markus, G. B. (1975). Birth order and intellectual development. *Psychological Review*, *82*(1), 74–88. https://doi.org/10.1037/h0076229

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics, 124*, 342–349.

Appendices

Chapter 2 Appendices (Long et al., 2019)

Appendix A: Infant recording information

Table 1. Infant demographics

Infant	Gender	Birth order	Maternal education	Ethnicity	Home language
1	Female	1	PhD	White/Caucasian	English
2	Female	1	Some graduate school	White/Caucasian	English
3	Female	1	PhD	White/Caucasian	English, Ukrainian
4	Male	3	Some college	White/Caucasian	English
5	Male	2	BA	White/Caucasian	English
6	Male	3	Some college	White/Caucasian	English

Table 2. Infant ages in recordings used for stimulus selection

Infant ages in recordings used for stimulus selection. Imitations per minute offers perspective on possible individual differences in rate of imitation.

Infant	Gender	Recording age of infant			Imitations per minute
1	F	3 mo 1 wk 4 dy	6 mo 0 wk 6 dy	9 mo 4 wk 1 dy	0.36
		3 mo 1 wk 4 dy	6 mo 3 wk 3 dy	9 mo 4 wk 1 dy	
2	F	4 mo 0 wk 2 dy	6 mo 0 wk 3 dy	11 mo 3 wk 2 dy	0.22
		4 mo 1 wk 2 dy	7 mo 1 wk 0 dy	11 mo 3 wk 2 dy	
3	F	3 mo 0 wk 4 dy	5 mo 0 wk 4 dy	10 mo 1wk 6 dy	0.25
		3 mo 0 wk 4 dy	6 mo 0 wk 4 dy	10 mo 1 wk 6 dy	
4	M	3 mo 2 wk 5 dy	6 mo 0 wk 3 dy	9 mo 3 wk 6 dy	0.02
		3 mo 2 wk 6 dy	6 mo 3 wk 6 dy	9 mo 3 wk 6 dy	
5	M	4 mo 2 wk 2 dy	6 mo 0 wk 4 dy	11 mo 2 wk 1 dy	0.02
		4 mo 2 wk 2 dy	7 mo 3 wk 1 dy	11 mo 2 wk 1 dy	
6	M	3 mo 2 wk 0 dy	5 mo 0 wk 2 dy	10 mo 0 wk 6 dy	0.13
		3 mo 2 wk 0 dy	6 mo 0 wk 2 dy	10 mo 0 wk 6 dy	
Average		**3 mo 2 wk 3 dy**	**6 mo 1 wk 3 dy**	**10 mo 2 wk 4 dy**	**0.16**

Chapter 2 Appendices (Long et al., 2019)

Appendix B: Stimulus Pair Selection

A number of labels were used heuristically during stimulus selection, but the experiment did not utilize these category labels to designate any aspect of imitativeness. Instead the study with the 18 listeners addressed a continuum of imitativeness only. There was only a preliminary attempt to match the number of selected items in the no, low, and high imitation groups, and we did not view such matching as important since the focus was a single continuum rather than categories. A visualization of this selection process is shown in Figure 1.

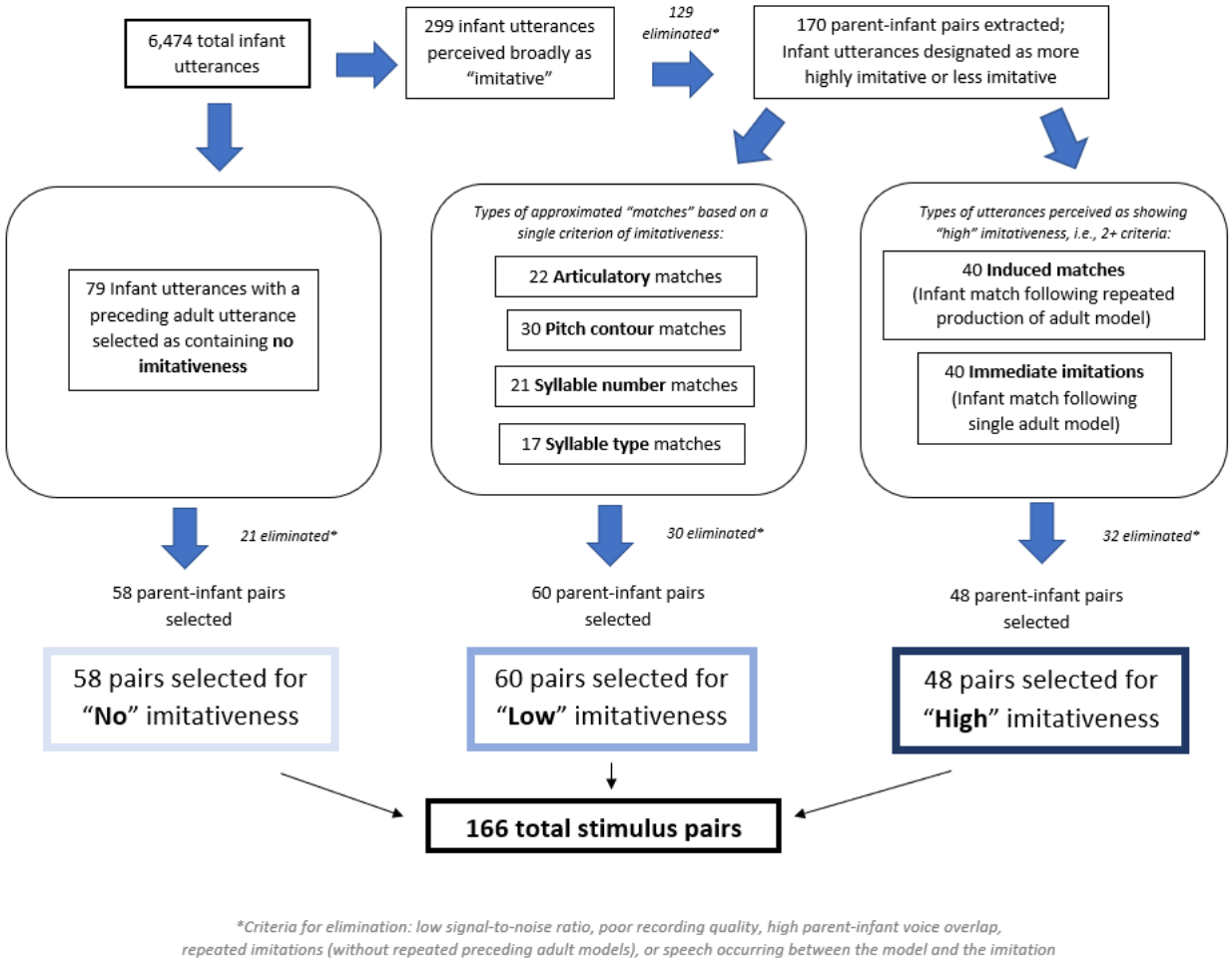

Figure 1. Visualization of selection process for stimulus pairs

Chapter 2 Appendices (Long et al., 2019)

Appendix C: Rating Scale

Figure 2 provides a screen shot of the continuous rating scale presented to listeners for making judgments on the degree of infant imitation. Listeners selected "Play" to hear a stimulus pair, then selected a position somewhere along the scale to rate *how* imitative the infant vocalization was compared to the adult model. Listeners pressed "Next" to continue and completed the task after 830 total ratings (166 stimulus pairs, 5 randomized blocks).

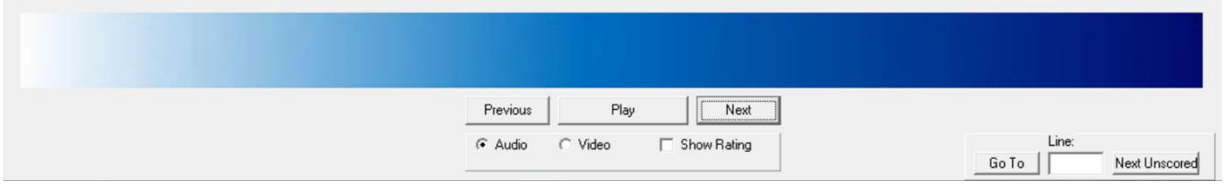

Figure 2. Rating scale
Continuous rating scale presented to listeners for making judgments on the degree of infant imitation.

To estimate the variability in individual stimulus pair ratings, we computed the mean rating (individual rater means, IRMs) across the 5 trials on each stimulus pair for each listener. We then calculated the stimulus pair means (SPMs) for ratings of each stimulus pair, that is, the means of the IRMs across the 18 raters. We similarly calculated the stimulus pair standard deviations (SPSDs). Figure 3 presents the SPMs versus the SPSDs, thus characterizing the consistency across trial judgments for each of the 166 pairs, aggregating the ratings from all 18 listeners. The parabolic shape of the distribution suggests that listeners were consistent in their judgments of very low and very high degrees of imitativeness but had greater variability in rating items for moderate levels of imitativeness. In other words, the consistency of judgments was not uniform across the range of trials and was greater for extreme judgments of "not imitative" and "highly imitative."

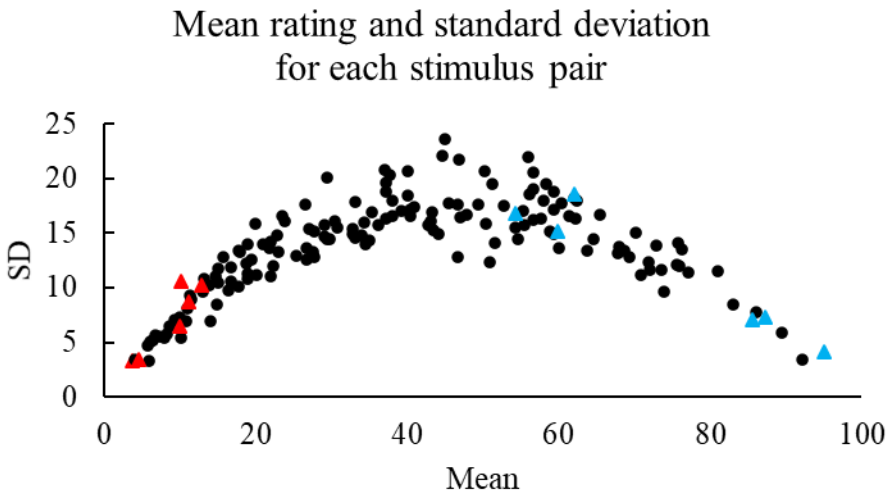

Figure 3. Rating scale usage and variation
Dots and triangles represent the average of the means and standard deviations across all 18 raters on each individual stimulus pair (N = 166). The data show listeners used more consistent ratings for extremely low and high degrees of imitativeness. ● = Stimulus pair not among the calibration items; ▲ = Low imitativeness calibration item, ▲= High imitativeness calibration item.

Ratings for the 12 calibration stimulus pairs are represented as red and blue triangles—low and high imitativeness, respectively—in Figure 3. These pairs had been selected by the first author and explicitly presented to listeners prior to the judgment task as examples of very low and very high degrees of imitativeness. The listeners consistently rated the low calibration pairs as having a low degree of imitativeness (M = 8.78, SD = 7.08), whereas the high calibration pairs were rated with greater variability (M = 74.12, SD = 11.42). Rater 1 rated all the low calibration pairs $\leq$ 5, and all the high calibration pairs $\geq$ 80.

An analysis of overall rating bias was calculated on the frequencies of individual rating values across the 0-100 scale as seen in Figure 4 (grouping 90-100 included 11 values; all other groupings included 10 values, i.e. there were 101 possible rating values in the scale from 0-100). With 18 raters and 5 stimulus-pair trial blocks of 166 items, there were a total of 14,940 ratings for the entire experiment. Lower rating judgments were used more often, suggesting a tendency to judge the infant utterances as having a low degree of imitativeness. Specifically, the total

number of ratings from 0 to 9 made up 29.0% of the total of all the ratings, whereas each of the other rating intervals made up on average 7.9% of the total.

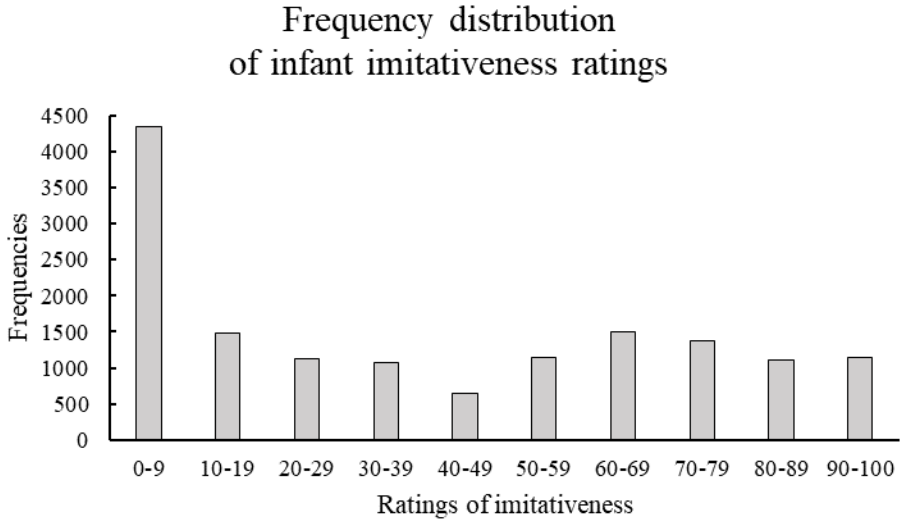

Figure 4. Frequency distribution of rating scale usage
Frequency distribution of the 14,940 ratings (166 stimuli x 18 listeners x 5 trials) used across the 0-100 scale. Listeners predominantly rated utterances as having a low degree of imitativeness (0-9).

Mean ratings of each listener across the five trial blocks were calculated to examine individual biases regarding degree of rated imitativeness, as displayed in Figure 5. The average rating of individual listeners was 39.3 (range: 16.2-55.2). Listeners consistently rated pairs as having a relatively low degree of imitativeness; all but three raters had an average rating below 50. The figure shows that the listeners significantly differed in rating bias (or criterion). These differences are reflected in the means and 95% CIs. Note in particular Rater 11, who shifted from a first trial mean rating of 17.9 to a fifth trial mean of 45.6. This suggests she changed her criterion or rating bias substantially across the trials. On the other hand, Raters 8, 9, 12, 16, and 2 scarcely changed their rating criteria across the five trials.

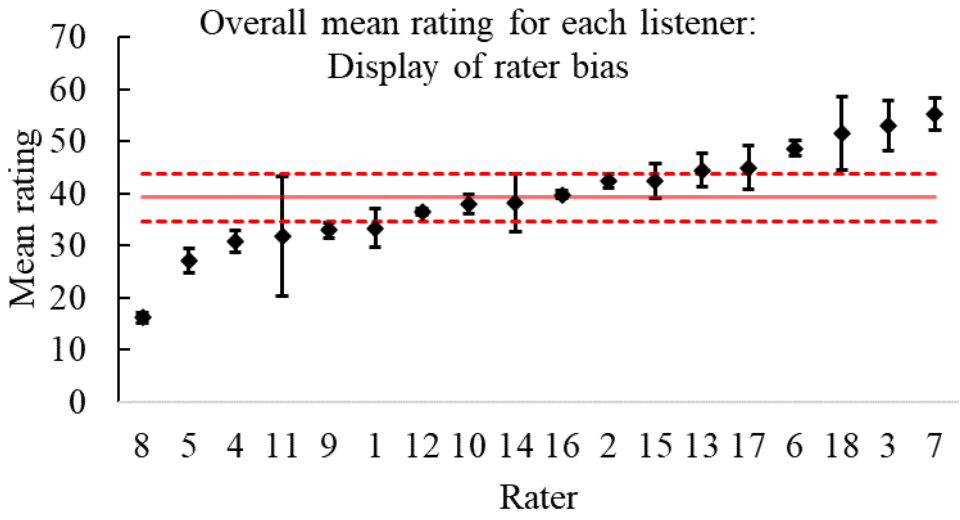

Figure 5. Display of mean individual rater bias (an intra-rater analysis)
Mean ratings for each listener, ordered from lowest (M = 16.2) to highest (M = 55.2), with 95% confidence intervals represented for each. Y-axis reflects range of rating scale, 0-100. The overall mean rating was 39.3 (95% CI = 34.7 – 43.8).

Evaluating rater bias differences between listeners, we compared each rater with all others on their mean ratings across the 166 pairs. Paired t-tests were calculated to compare IRMs across the 18 raters. Specifically, the IRM for each rater (N = 18) was compared to the IRMs for all other raters, yielding a total of 153 possible paired comparisons t-tests (n=166) as seen in Table 3. 130 out of the 153 comparisons were found to be significantly different ($p < .05$), suggesting raters were making judgments the means of which were systematically different from those of other raters, that is, that the raters showed different rating biases. In other words, 85% of the comparisons showed strong differences in ratings between listeners. A 2x2 chi-square test of independence supports the idea that listeners were systematically different from each other in their perceptions of the degree of imitativeness in stimulus pairs, $\chi^2(17) = 101.69$, $p < .001$. It is important to emphasize, however, that the bias differences between raters are independent of the correlations that obtained among raters. Even though the bias differences were very discernible and statistically significant, it is also true that the raters showed strong agreement in terms of correlations of their ratings with each other.

Table 3. Rating bias across stimuli between raters (an inter-rater analysis)

130 out of 153 comparisons (85%) were found to be significantly different ($p < .05$), suggesting raters were making judgments that were systematically different from each other in terms of bias. Thus Rater 1's mean judgments on the 166 stimuli were statistically different from those of Raters 2, 3, 5-8, 10, and 13-18 (either higher or lower in each case).

Rater	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	<.001	<.001	0.096	<.001	<.001	<.001	<.001	0.778	0.030	0.358	0.058	<.001	0.005	<.001	0.001	<.001	<.001
2		<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.036	<.001	<.001	0.209	0.017	0.962	0.158	0.099	<.001
3			<.001	<.001	0.019	0.095	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.412
4				0.016	<.001	<.001	<.001	0.180	0.001	0.559	0.001	<.001	<.001	<.001	<.001	<.001	<.001
5					<.001	<.001	<.001	0.001	<.001	0.012	<.001	<.001	<.001	<.001	<.001	<.001	<.001
6						<.001	<.001	<.001	<.001	<.001	<.001	0.010	<.001	<.001	<.001	0.012	0.078
7							<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.015
8								<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
9									0.006	0.452	0.008	<.001	0.001	<.001	<.001	<.001	<.001
10										0.001	0.391	<.001	0.917	0.010	0.210	<.001	<.001
11											0.002	<.001	<.001	<.001	<.001	<.001	<.001
12												<.001	0.297	<.001	0.026	<.001	<.001
13													<.001	0.160	0.001	0.696	<.001
14														0.012	0.279	<.001	<.001
15															0.095	0.080	<.001
16																0.001	<.001
17																	<.001

Chapter 2 Appendices (Long et al., 2019)

Appendix D: Audio Wave Files

Table 4. Audio wave file means and standard deviations

Audio wave file information. Raw rating means and SDs of individual audio files across all raters and judgments.

File	Mean Rating	SD
Audio 1.WAV	3.90	3.40
Audio 2.WAV	6.13	3.40
Audio 3.WAV	25.01	13.36
Audio 4.WAV	23.13	13.04
Audio 5.WAV	50.34	14.41
Audio 6.WAV	51.57	18.21
Audio 7.WAV	75.39	12.03
Audio 8.WAV	72.13	11.73
Audio 9.WAV	93.40	6.31
Audio 10.WAV	86.25	8.30

Audio files also available at:
https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01340/full#supplementary-material

Chapter 3 Appendices (Long et al., 2020)

Appendix E: Focus of Prior Literature in Infant Vocalizations

It has been our impression that most literature in infant and child speech and language development has tended to gather data in interactive circumstances. The work has tended to place primary emphasis on the vocalizations of babies in terms of their vocal interactivity and on contingencies between adult vocalizations and infant responses as well as on adult elicitations and responsivity to infant sounds. These tendencies in the literature, we have surmised, have yielded relatively little attention to endogenously produced infant protophones. Assuming our impression is correct, the literature's tendency is surprising since our data in the present paper suggest most infant protophones are not directed to other persons. But is our impression of the literature consistent with the facts?

In response to a reviewer suggestion we conducted a PubMed search. We focused on abstracts only. The term "infant vocalization" returned many abstracts that showed no emphasis on social vs. endogenous human infant vocalization, and so were irrelevant to our impression of a primarily social emphasis in the literature. Some abstracts that were returned in the search, for example, merely reported acoustic data on infant sounds, with no mention of either independent/endogenous or interactive/social production. Many were about cry only, not protophones. A great many were not about human vocalization at all (birds were a particular focus). We ignored all such articles as well as articles from or in collaboration with our lab (i.e., with Oller as at least a co-author).

We examined the abstracts for the first 160 articles returned by the search (dates of the 160 articles ranged from 2014 to 2020); only 18 of them could be judged with moderate certainty based on the abstracts regarding having an emphasis on social use of human speech-like

vocalizations as opposed to endogenous or independent use. None of these 18 appeared to have attempted to address the question of the current paper (actually counting and focusing on a comparison of rates of social and endogenous use of infant vocalization). Also none seemed to have placed primary emphasis on independent, endogenous production. Fifteen revealed a focus on the social-interactive use of infant sounds, while 3 described use of infant sounds when infants were not interacting, though consideration of interaction was not excluded in these cases. So all in all, the review seemed to suggest our impression that a lack of emphasis on endogenous protophone production in the literature is essentially accurate.

Chapter 3 Appendices (Long et al., 2020)

Appendix F: Opinion Survey on The Function of Infant Vocalizations

As background information for the primary goal of this research, we sought survey data where both parents and non-parents were asked to provide estimates of how often they thought infants vocalize with social directivity and without social directivity based solely on a reflection of their own experiences around infants. We hypothesized that survey participants would provide evidence supporting our general impression of the literature on vocal development, an impression suggesting that socially-directed vocalization is emphasized more often than endogenous vocalization.

Materials and Methods

We collected survey data using Amazon Mechanical Turk ("mTurk") to provide a perspective on the observational data in the main text of the article and an empirical evaluation of the suspicion that not only many researchers in child development, but also the general public, have the impression that infants predominantly vocalize socially. mTurk is increasingly used as an online recruitment tool for participation in experimental studies and academic surveys as a quick method to obtain many responses from the general public. mTurk has been shown to be slightly more representative of the US population than of other countries and is considered to be as reliable as traditional survey methods (Buhrmester et al., 2011; Hauser & Schwarz, 2016; Paolacci & Chandler, 2014). mTurk qualifications used for this study included: 1) having a HIT Approval Rate greater than 95%, and 2) at least 50 Approved HITs. These ratings ensured that all participants were experienced and had been deemed acceptable participants in prior mTurk studies. Such qualifying indicators are regularly used by mTurk researchers to safeguard against inaccurate and inattentive workers.

Survey instructions

Following consent, participants were presented the following written instructions for the survey:

> *This is a study evaluating your perception of how often babies make different kinds of sounds and why they make them. You will be asked to consider sounds produced by babies at three different ages: Infants who are 3-months, 6-months, and 10-months old. Across any given day, consider all the sounds (or "vocalizations") babies make. Your task is to estimate the percent of these sounds that serve a particular function (social or endogenous). In answering the questions, consider your previous experiences (if any) around babies and give an intuitive guess for each question. When thinking about your responses, only consider babies who are typically developing, not those who may have special conditions causing atypical development. You are not expected to be an expert on this, and there are no wrong answers. You will be asked to give an intuitive response. Your responses will be required to sum to 100 (e.g., 100%).*

Participants estimated a percentage of social and endogenous infant vocalization functions at three ages (3-month-olds, 6-month-olds, and 10-month-olds) for a total of six judgments. Means and standard deviations of these responses were calculated. The data in the main text based on laboratory-coded observations of real infants in audio-video recordings were collapsed to yield the same categories used in this opinion study (social vs. endogenous) for each infant utterance.

Survey participants

300 participants completed the online survey, and 239 participants' data were used in final analysis based on adequate responses to three attention checks distributed throughout the survey. The attention checks ensured that the responders were not robots and that the responders were sufficiently knowledgeable in English to have understood the questions clearly. The attention checks were questions presented to all participants:

1) *Provide a word that means the opposite of happy.*
2) *Select the option that includes 5 times a week:*
    a. *one time a week,*

b. two to three times a week,

c. four to five times a week,

d. six to seven times a week.

3) Type in the number sixty.

For the first attention check, a variety of English words meaning “not happy” were accepted. For the second attention check, options C and D were accepted. For the third attention check, only the number “60” was accepted. Failure on two of the attention checks resulted in rejection of the participant. In general, such questions capture robots, which fail usually to answer the questions in a meaningful way. For language background, the participants were asked to list the languages they speak in the order of most to least fluent. Only individuals indicating that English was at least second on their list were included. An additional measure to try to limit the group to English speakers was the inclusion of a worker qualification in the mTurk survey settings that required the computer system location to be in one of the following countries where English is the primary spoken language: AU, CA, NZ, GB, or US. In other words, the worker had to reside in or be taking the survey on a computer registered in one of these countries. Detailed demographics of the mTurk survey participants are presented in Table 5.

Table 5. mTurk survey participant demographics
Participant demographics for opinion study.

Age		Gender		Education		Number of children		Frequency around children	
18-21	3	**Male**	139	**Less than HS**	2	**None**	124	**Never**	29
21-34	126	**Female**	97	**HS/GED**	29	**1**	41	**Rarely**	83
35-44	50	**Other**	3	**Some college**	48	**2**	41	**Sometimes**	62
45-54	34			**Associate's**	33	**3**	21	**Frequently**	46
55-64	24			**Bachelor's**	111	**4+**	12	**All the time**	19
65+	2			**Master's**	9				
				Doctorate (PhD)	2				
				Professional Degree (JD, MD)	5				

Results

Figure 6 shows the survey participants' distribution of responses on relative percentages of protophones across the three ages. On average across the three ages, the respondents thought approximately 43% of infant protophones were endogenous. In addition, they thought infants produce fewer endogenous vocalizations at the end of the first year (36%) than at the beginning (50%). Thus, the respondents believed more than half of infant protophones are socially directed and many more than half by 10 months.

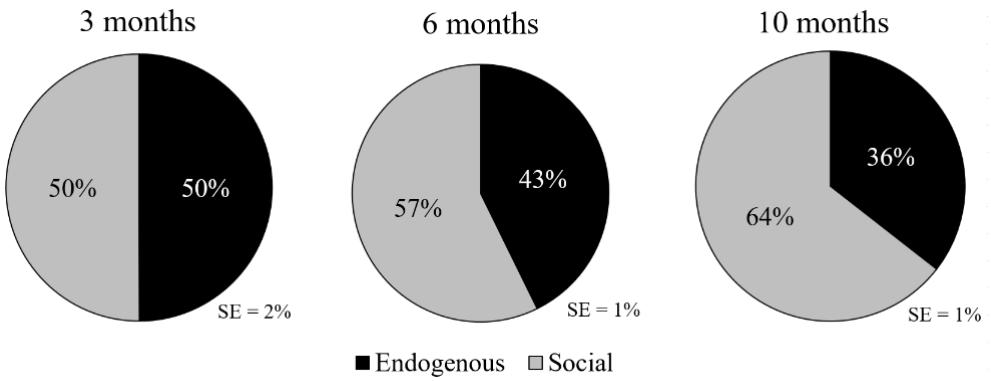

Figure 6. mTurk opinion study on social directivity of infant protophones across 3 ages
Opinions of the survey participants on how often infants use protophones socially and endogenously. Participants believed infants decrease the percentage of endogenous protophones between 3-10 months, from 50% at the youngest age to 36% by the oldest age.

Parents and non-parents reported similar percentages of social and endogenous vocalizations. Overall, parents reported infants used protophones socially 58% of the time, whereas non-parents reported 57%. Males and females also estimated very similar percentages of social protophones (58 and 57% respectively). Persons who self-identified as being around kids "all the time" estimated that infants produce 58% social protophones, while those who self-identified as never being around kids estimated 55%. For all these comparisons (parents v. non-parents, males v. females, always around kids v. never around kids), the estimated percentage of social protophones was higher at 6 than 3 months and higher at 10 than 6 months.

Chapter 3 Appendices (Long et al., 2020)

Appendix G: Expected and Actual Circumstance Durations

During recording, parents were asked to participate in two recording protocols. For twenty minutes each, parents were asked to engage in face-to-face interaction with their infant (parent seeking to be Engaged with the infant) and to converse with an interviewer while the infant was in the room (parent and infant Independent). Each protocol was initially labeled per the "expected" session protocol. These expected protocol durations are presented in Table 6. The times varied because infant state varied, and sessions were often readjusted for length to keep infants comfortable and often because parents requested the readjustments.

Table 6. Expected protocol durations
Duration of expected protocol sessions in Engaged (Engd) and Independent (Ind) protocols for each infant at each age. The minimum duration was 12:08, maximum duration 22:03, with an average duration of 19:10.

Infant	Gender	Length of recording					
		Engd	Ind	Engd	Ind	Engd	Ind
1	F	00:19:41	00:14:13	00:18:42	00:19:29	00:19:58	00:19:58
2	M	00:20:03	00:20:19	00:20:47	00:21:05	00:21:16	00:20:24
3	M	00:22:03	00:22:01	00:20:20	00:20:13	00:20:11	00:12:52
4	F	00:19:01	00:19:39	00:16:03	00:19:37	00:19:51	00:19:52
5	M	00:16:11	00:19:51	00:20:54	00:18:11	00:20:52	00:20:54
6	F	00:19:48	00:19:53	00:12:08	00:14:23	00:19:08	00:19:54
Mean age		**3 months**		**6 months**		**10 months**	

To encourage naturalistic interaction throughout all protocols, parents were not restricted from engaging with the infant or another person if warranted despite the expected protocol (e.g., to comfort the infant if crying during the *Independent* protocol or to answer a question from a staff member during the *Engaged* protocol). Because the parent would occasionally engage with the infant for notable periods of time during the "expected" *Independent* protocol, or to converse with a staff member during the "expected" Engaged protocol, each recording was re-coded, segmenting it into "actual" *Engaged* and *Independent* circumstances. These periods of time were

summed at each age for each infant to create actual protocol durations, shown in Table 7. Four cells in the Table (for Infants 1 and 6 at three and six months in *Independent* circumstance) showed actual protocol durations of less than 5 minutes each, highlighted with an (*).

Table 7. Actual protocol durations
Duration of actual segments concatenated with Engaged (Engd) and Independent (Ind) activity for each infant at each age. Overall, there were longer periods of time in the Engaged circumstance compared to the Independent circumstance. The minimum duration was 00:58, maximum duration 32:52, with an average duration of 19:06.

Infant	Gender	Length of recording					
		Engd	Ind	Engd	Ind	Engd	Ind
1	F	00:32:38	00:01:16*	00:33:48	00:04:23*	00:20:34	00:19:22
2	M	00:27:59	00:12:24	00:26:59	00:14:53	00:23:34	00:18:08
3	M	00:22:46	00:21:19	00:23:08	00:17:28	00:25:35	00:07:29
4	F	00:23:26	00:15:15	00:10:31	00:25:08	00:24:27	00:15:16
5	M	00:22:00	00:14:02	00:20:54	00:18:11	00:21:45	00:19:55
6	F	00:35:52	00:01:37*	00:25:33	00:00:58*	00:24:02	00:15:00
Mean age		**3 months**		**6 months**		**10 months**	

A ratio of expected over actual times for each circumstance and age is presented in Table 8. For most circumstances, larger ratios are seen in *Engaged* circumstances, showing parents were often inclined to engage with their infant in both expected *Engaged* and expected *Independent* circumstances, often running counter to the protocol instructions.

Table 8. Ratio of expected over actual protocol durations
Ratios for actual over expected protocol durations in Engaged (Engd) and Independent (Ind) circumstances. Larger ratios are seen in the Engaged circumstances for all but two infant ages (Infants 4 and 5 at six months).

Infant	Gender	Length of recording					
		Engd	Ind	Engd	Ind	Engd	Ind
1	F	1.66	0.09	1.81	0.23	1.03	0.97
2	M	1.40	0.61	1.30	0.71	1.11	0.89
3	M	1.03	0.97	1.14	0.87	1.27	0.58
4	F	1.23	0.78	0.66	1.28	1.23	0.77
5	M	1.36	0.71	1.00	1.00	1.04	0.95
6	F	1.81	0.08	2.11	0.07	1.26	0.75
Mean age		**3 months**		**6 months**		**10 months**	

Chapter 3 Appendices (Long et al., 2020)

Appendix H: The Origin of Vocal Flexibility in Humans and the Fitness Signaling Hypothesis

Oller and various colleagues, including Long and Bowman (and especially Ulrike Griebel), have written elsewhere on the idea that human development provides key information about likely sources of the selection pressures that have driven hominins to differentiate dramatically from our ape cousins in vocal communication (Griebel & Oller, 2008; Oller et al., 2016; Oller & Griebel, 2005, 2008). We largely share this reasoning with J. L. Locke who formulated a similar proposal independently (Locke, 2006, 2009). In this evolutionary developmental biology or "evo-devo" framework (Bertossa, 2011; Carroll, 2005; Müller & Newman, 2003; Newman, 2016) we have formulated a natural logic of development and evolution, where it is proposed that foundational communicative capabilities must develop in order for subsequent capabilities (ultimately required for language) to be possible. Within that reasoning, an essential foundation for language evolution and human linguistic development is a flexible system of expression, where *all* the elements (the vocal modality has proven to be preferred) can be produced with *any* illocutionary intent (any function). One of those possible intents had to have been exploration of vocalization itself, for no social purpose. Another would have been emotional expression, whether of positive or negative emotions. And another, of course, would have been (or would have developed quickly to become) social interaction involving sharing of emotional states and information (going beyond purely manipulative functions such as pleading for help, a kind of function that is common in mammals). Crucially we have posited that human vocal social interaction is itself founded in flexible vocalization. According to the reasoning, one cannot flexibly share states and information vocally, but can

only engage in manipulative vocal interactions (threatening, courting, soliciting…, the kinds of vocal interactions seen in mammals in general), unless one has the flexibility to express states that are *not* bound to particular manipulative goals.

The human infant appears to have such a vocal capability from birth (Jhang & Oller, 2017; Oller et al., 2019), producing ~3500 protophones daily (Oller et al., 2019). But the other apes appear to have no such capability. In 1700 minutes of longitudinal observation of 3 bonobo infants with their mothers we found not a single instance of a "protophone-like" sound produced by a bonobo infant that was interpreted by the coders as "exploratory" or "playful" (Oller et al., 2019). All the "protophone-like" sounds produced by the bonobo infants that could be interpreted for function/affect were interpreted as negative/complaint or plea-like vocalizations (the infant seeking help from the mother or simply complaining).

Importantly we also found not a single case of a maternal vocalization directed to one of the bonobo infants. The mothers were very responsive to the infant pleas, but never vocally. It appears chimpanzees are similarly constrained in vocal interaction with infants (Kojima, 2003). The evidence is consistent with the proposed natural logic: In the absence of flexible vocalization on the part of the infant, there is no basis for development (or evolution) of flexible vocal interaction.

So the fundamental question becomes, what selection pressures could have resulted in a flexible vocalization capability before language existed, indeed before vocal social interaction in apes (not obligatorily manipulative in any particular way or limited to a specific single goal) existed? The results of selection pressures had to have been advantageous to the individuals subject to those pressures at the time they first appeared. Thus they could not have been selected as preparation "for language" or language-like communication because language or language-

like communication did not yet exist. This is where the fitness signaling idea has traction. Hominin infants, who were more altricial than other ape infants, were more in need of parental care and for a longer period of development than other ape infants. In accord with the proposal, hominin infants were thus under heightened selection pressure to signal their wellness, and vocalization became one of the targeted means of doing that.

Hominin infants were, then, selected to produce protophone-like sounds endogenously and flexibly, especially in circumstances of comfort and lack of immediate need, because in that way, caregivers could recognize and judge the wellness of the infants. The advantage to the caregivers was greater efficiency in their investments in offspring, yielding presumably more numerous progeny in subsequent generations. Hominin infants are thus seen as having been in competition with each other for parental investment and so were selected generation after generation to be increasingly inclined to vocalize in a variety of states including in comfort and with illocutionary flexibility, that is, to produce protophones. The availability of these flexible sounds, recognized by caregivers (who were themselves, according to the proposal, under selection pressure to accurately recognize the well-being of their infants) afforded the opportunity for comfortable vocal interaction among parents and infants (a type of interaction largely absent in other apes, with the exception of certain "close calls" that in some cases occur during grooming), during which parents had further opportunity to observe and even elicit vocal fitness signaling. The flexible vocalizations of the ancient hominin infants provided the raw material of vocalization for parent-infant vocal interactions. The parents of the infants, having been the beneficiaries of the same selection pressure on vocal flexibility from their own infancies, are imagined within the proposal to have developed further vocal flexibility as they

matured, along with increasing interest in observation of their infants' vocal capacities as information about their fitness.

Bonding and attachment of hominin parents and infants seem to have come to be pursued in part during and through these flexible and comfortable vocal interactions. Generation by generation the infant tendency to vocalize freely and the parental tendency to intuitively recognize the import of the protophones grew, according to the proposal, cyclically, ratcheting up the vocal capacities and vocal interactions of hominins across the life span and forming foundations for additional vocal communicative growth.

The fitness signaling function of the protophones did not require that the sounds be intended by the infants as fitness signals—the perlocutionary effect, that is the reaction of the parent in interpretation of the protophones *as* fitness signals needs to be distinguished from the illocutionary intent of the infant in producing the protophones. The infant's intent had to be variable on different occasions (or the vocal capacity would not have been functionally flexible), and crucially, at least some of the time that intent had to be purely exploratory, the infant expressing interest in the sound production itself, while on other occasions the same sounds had to be producible as expressions of varying emotional states or in seeking to engage or maintain interactive engagement with a caregiver.

The reason the parent's reaction needs to be distinguished from the infant's intents, is in part that regardless of the infant's intent with protophone production, the parent's interpretation would affect the parent's decisions about investment. The parent's interpretation of the infant's fitness would have resulted from the protophone production (and of course many other signs of fitness of the infant including body movement, skin color, eye contact…) even if the infant had not intended the sounds as fitness signals. Of course there is no reason to doubt the infant could

at least on some occasions produce protophones indeed for the purpose of soliciting parental social attention, and in that way may have intentionally been seeking investment. But not always, and that is the key point. The hominin break from the ape background depended, according to our reasoning, on selection for infants who had both the capability and the inclination to produce all the protophone types (on different occasions) in *any* state and with *any* intention. We have striven to emphasize in all our writings about this point that language in all its forms requires this kind of flexibility of expression, as revealed by the fact that every word or sentence of any language can be produced in any circumstance of state or intention. Put another way, humans can utter any word or sentence with any chosen illocutionary force. The fact that human infants from the first month of life appear to be able to do the same with protophones suggests to us that a fundamental break from the ape background occurred when hominin infants were selected to produce protophones in any state and with any purpose. We reason that parental selection of infants based on their interpretation of protophones as fitness signals resulted, generation after generation, in infants inclined to produce such sounds more and more copiously.

Importantly, the proposal does not suggest that the selection pressures on this system of infant endogenous protophone production and parental interest in those sounds and elicitation of them would have abated in modern times. It is an empirical question how and to what extent the pressures apply nowadays (we are already engaged in studies of behavioral responses of parents and other adults listening to infant sounds and are planning physiological studies of adult responses as well).

In the societies we ourselves have studied, parental attention to infants in the first year tends to be intense, although it varies, for example, by socio-economic status. But even in the circumstance where parents are involved relatively little in interaction with their infants, we have

never observed a human infant who did not produce massive numbers of protophones. A key unresolved empirical matter concerns the extent to which modern infants produce protophones in societies where, for example, infant mortality is high, and where there is parental resistance to interacting vocally with very young infants. Resistance even to naming infants until they have proven their survivability has been invoked as a possible corollary of such resistance to vocal interaction with infants in at least some societies. We know of no direct studies of protophone rates produced by infants in such societies, although there are empirical reports suggesting much reduced levels of IDS (see (Cristia et al., 2019)).

Our rationale, then, is built on our proposal that more attention in human development research needs to be directed to the endogenously-produced protophones. That the majority of them seem to be produced without social directivity is surprising to us, and we presume it will be surprising to most readers.

Our methods depend on coders' acting as intuitive observers, noticing moment by moment the direction of infant attention, and taking stock of the fact that infants often direct their attention away from interaction even during periods where the parent is eliciting it, and where the infant is intermittently participating actively in it. As noted by Maya Gratier in her review of a previous version of this work, one should not underestimate the potential importance of those occasions (even if they are relatively rare) where the infant is indeed fully engaged and vocalizes in harmony with the caregiver, that is, where the active infant applies its endogenous capacities in directed dyadic communication. Indeed there is reason to suspect that those events of very engaged face-to-face vocal interaction are critical in social and language development. The present paper emphasizes that the infant's endogenous vocalization provides raw material that is required for such active, comfortable vocal interaction.

Furthermore the endogenous vocal tendencies of infants appear to play a significant role in the development of the vocal capacity itself—vocal exploration may serve as a sort of practice in phonatory and articulatory skills that not only provide fitness signals but at the same time lay groundwork for subsequent vocal expression. Note again, that the initial selection pressures that have driven the production of protophones in hominins would have had to involve advantages applying *before* linguistic communication existed. Consequently it seems selection pressure on a practice function could not have operated in isolation but would have been logically dependent upon other selection pressures to establish primitive flexible communication through vocalization in the absence of a language target. The fitness signaling function appears to provide a selection pressure that could have operated in the absence of modern language or even of primitive protolanguage.

We are unaware of other proposals that could explain the initial break with the vocal communication limitations of our ape ancestors (although our own proposal is shared by J. L. Locke). It has been suggested we consider alternative proposals regarding the origin of vocal language such as those of Falk (2004) and Dissanayake (1992). There are quite a few psychologists, linguists, biologists, and cognitive scientists that we could add to such a list (e.g., (Deacon, 1997; Fitch, 2000; Gärdenfors, 2004; Hurford, 2011; Sinha, 2004)) But as far as we know, none of these proposals offers an explanation for the initial break from the ape background with regard to vocal flexibility, the event that we think is a prerequisite to all the other requirements that would have had to evolve for language to have ultimately emerged (infant-directed speech, vocal imitation, learned vocal performatives, primitive syntax, and so on).

A proposal of Robin Dunbar is perhaps the most closely related to our own (Dunbar, 1993, 1996, 2004). He has argued that vocalization in ancient hominins may have assumed a role similar to that of grooming as hominin group sizes increased and there was insufficient time in the day to physically groom all the necessary members of the group. “Vocal grooming” (which was posited even earlier by Morris (1967) could service multiple members of the group simultaneously. The grooming function was thought to provide a platform for elaboration of human vocalization in subsequent evolution. That close calls (and lip smacks, see (Locke, 2008)) occur sometimes in primate grooming suggests there may have existed a comfortable social function for some vocalizations in our distant ancestors. In addition, protophones, produced in interactive circumstances, can be thought of as a kind of vocal grooming. But the Dunbar proposal does not incorporate the evo-devo perspective, wherein it is assumed that new vocal capacities would have likely been selected for first in infants, whose subsequent development could have laid the groundwork for the occurrence of even more elaborate vocalizations in grooming adults and in other kinds of interactions among adults.

In our opinion, the grooming hypothesis of Dunbar also requires that the earlier question be answered: How might infant vocalizations with functional flexibility (including the flexibility to have been used in grooming) have been selected for before language existed or before other elaborate forms of vocal interaction existed? We propose that fitness signaling offered the opportunity for selection of infants with greater inclination to vocalize flexibly, and once that greater inclination was in place, an important consequence could have been the development of capabilities yielding social grooming and other functions in infancy and in adulthood.

Our theoretical inclination is based 1) on the proposed natural logic of how a language capability could in principle evolve (vocal flexibility is required for all the features of vocal

language and language use) (Oller et al., 2016), 2) on a common tenet of evo-devo, wherein it is assumed and observed that natural selection tends to target development; if a new structure or capability is to emerge, its genetic foundations must be targeted; minor genetic changes can produce significant changes in structures or capabilities through epigenetic, self-organizational development, and 3) on the fact that protophones have been observed to occur copiously in all human infants long before language and in fact months before many of the presumed prerequisites to language. Thus we propose minor genetic changes in ancient hominins could have resulted in greater flexible vocal activity in hominin infants, and that that greater activity could have had cascading consequences on later capabilities relevant to vocal communication.

Chapter 4 Appendix (Long et al., in submission)

Appendix I: Considerations Regarding Infant-Directed Speech

The literature on early language suggests infant-directed speech (IDS) may also influence the emergence of canonical babbling, as previous research has highlighted the effects of social interaction on babbling (Albert et al., 2018; Goldstein et al., 2003; Goldstein & Schwade, 2008) and conversely, the effects of babbling on caregiver speech during interaction (Elmlinger et al., 2019). During our analyses, we ran a secondary main effects model including IDS as a variable. Our coding protocol also included counts of both infant- and other-directed speech (i.e., speech between two adults) in each segment, affording the opportunity to compare counts or proportions of IDS to CBRs. We found a significant effect of IDS on CBRs (p = .034, b = -.0004), but notably, this effect was extremely small and negative. Furthermore, the correlation between canonical babbling ratios and total IDS showed a weak, negative correlation (r = -.02). Because we continue to believe IDS is a variable worth exploring further as an influence on canonical babbling—both as a continuous variable based on the coded amount of IDS or as a categorical factor (i.e., Low vs High IDS or No vs Any IDS) based on questionnaire judgments of the same type as explored for TT and VP in the present research—we plan to examine IDS effects on canonical babbling more explicitly in a future paper.