Social and non-social functions of infant vocalizations

Helen L. Long[1*], Dale Bowman[2], Hyunjoo Yoo[3], Megan M. Burkhardt-Reed[1], Edina R. Bene[1], D. Kimbrough Oller[1, 4, 5]

[1] Infant Vocalizations Laboratory, School of Communication Sciences and Disorders, University of Memphis, Memphis, Tennessee, United States of America

[2] Department of Mathematics, University of Memphis, Memphis, Tennessee, United States of America

[3] Department of Communicative Disorders, College of Arts & Sciences, The University of Alabama, Tuscaloosa, Alabama, United States of America

[4] Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee, United States of America

[5] Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria

* Corresponding author

Email: hlong@memphis.edu

HLL, DB, and DKO contributed to the design, analysis, interpretation, writing, and review of this work. HY, MMB-R, and ERB contributed to the interpretation and review of this work.

Competing Interests: The authors have declared that no competing interests exist.

Data Availability Statement: The recordings cannot be made publicly available because the conditions of IRB approval for making the recordings do not include permission from the parents to make the recordings publicly available. To provide the raw recordings would violate the confidentiality requirements of the data collection. We can supply the data worksheets including durations and counts obtained from coders. From these sheets all the calculations have been made that are included in the paper.

Abstract

Research on infant vocal development is focused primarily on vocal interaction with caregivers, where it appears to be largely assumed that infants vocalize mostly for the purpose of interaction. A survey of both parents and non-parents indicated that public opinion conformed to the expectation that infant vocalization is mostly socially interactive. However, we report that in laboratory recordings of infants and their parents, the bulk of infant speech-like vocalizations ("protophones") were directed toward no one, and instead appeared to be generated endogenously in exploration of vocal abilities. The tendency to 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, not as a passive recipient of vocal input.

Keywords: Speech development[1], Social interaction[2], Illocutionary force[3], Prelinguistic communication[4], Origin of language[5], Language development[6], Evolutionary-development[7]

1 Introduction

The study of vocal development has been dominated by the expectation that infants primarily vocalize in a speech-like manner when they are in social engagement, an expectation suggesting social interaction drives prelinguistic vocal development (1–6). Granted, social learning is required in order for infants to acquire the language-specific syllables and phonemic elements and the largely arbitrary pairings of words with meanings in languages. Thus, there can be no doubt that social interaction plays a critical role in infant vocal learning and language acquisition. Surprisingly, however, we know little about the extent to which infants actually engage in directed vocal interaction using the speech-like sounds or "protophones" of infancy (which include both canonical babbling and precanonical speech precursors in accord with the terminology of Oller, 2000), as opposed to simply vocalizing playfully or exploratorily. The proportion of infant protophones that are socially-directed has, to our knowledge, never been previously quantified, so the extent to which infant protophone production may be primarily endogenous rather than social is unknown.

Even so, infant vocalization, especially in the context of social interaction, has been researched for half a century (8–13). 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 (14–17). Experimental studies in the still-face paradigm (18) have shown that by 5-6 months of age, infants increase the rate of protophone production when the parent disengages from an ongoing vocal interaction (19,20), suggesting infants by that age seek to repair broken interactions with increased vocalization.

The long tradition of research in infant attachment and bonding (21–24) has included a distinct emphasis on the parent-infant dyad as the fundamental unit of human social and emotional development. Winnicott (6) 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 the idea has been taken too far, we think, being 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.

The low level of focus on the infant as an agent of vocal development in prior research might be in part an unintended consequence of the radical behaviorist tradition that for many decades treated behaviors as *responses* rather than *actions* (25,26). Panksepp and his colleagues have argued forcefully 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 (27–30).

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 (31–33). In the Supplementary Material to a published article based on recordings made in our own laboratory, we reported that infants across the first year of life produced the majority of their protophones when gaze was *not directed* toward another person (34). Also in a small-scale study 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 (35). 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" (p. 481) (36) in the absence of social interaction. 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 (37).

These findings suggest vocalizations are commonly produced with non-social functions. 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 indirect. We propose that this vocal exploration may have a deeply significant role in vocal development, alongside the support of caregiver interaction and ambient language exposure.

In spite of the possible importance of exploratory vocalizations in language development, to our knowledge there is no published evidence specifically targeting the social-directivity of infant protophones or the lack of it. As noted above, existing evidence about social-directivity of infant protophones is indirect. The necessary work requires considering gaze direction during infant vocalization and the extent to which infants may bid for attention vocally even when they are not in the same room with caregivers. It 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. Only with such work will it be possible to reliably quantify proportions of non-socially-directed infant protophones compared with rates of socially-directed ones.

Furthermore, we deem it important that such quantification be established *across contexts* in the first year of life. Prior studies suggest the proportions of non-socially directed 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 self-organization in prelinguistic vocal development (31,33) and may help establish perspective about relative roles of endogenous and interactive factors in vocal development.

Our approach to placing these issues in perspective not only takes stock of the vast literature on infant development, where endogenous vocalization has overwhelmingly taken a distant back seat to social interaction, but also considers the impressions of parents and potential parents obtained through a survey about the relative roles of endogenous and social vocalization in infant development. We compare the survey data with careful counts of recorded infant protophones both when they appear to be directed socially and when they appear to serve endogenous purposes of the infant.

1.1 Specific aims and hypothesis

Our primary goal is to determine the extent to which infants produce vocalizations in two ways: *With* and *without* social directivity at three ages across the first year of life, and in two circumstances: An *Interactive* circumstance, where the parent is instructed to interact with the infant, and a *Non-Interactive* circumstance, where the parent is present but engaged in a separate conversation with an adult. This quantification is hoped to provide a standard against which the traditional view of infant protophones as being predominantly a social phenomenon can be judged. As a precursor to the primary goal, we sought survey data where both parents and non-parents were asked to provide estimates of how often they thought infants vocalized with social directivity and without social directivity based solely on a reflection of their own experiences around infants. In this study, we hypothesize that:

1. **Part 1: Opinion survey on the function of infant vocalizations**. Survey participants will provide evidence supporting the general impression of the literature on vocal development, an impression suggesting that socially-directed vocalization is predominant, while non-socially-directed vocalization is relatively uncommon.
2. **Part 2: Observational study on the function of infant vocalizations**. In naturalistic laboratory recordings, infants will produce more non-socially-directed vocalizations and fewer socially-directed vocalizations, across two circumstances where parents are:
    a. instructed to interact with the infant (*Interactive*), and
    b. engaged in an interview with another adult (*Non-Interactive*).

2 Methods

2.1 Part 1: Opinion survey on the function of infant vocalizations

We collected survey data using Amazon Mechanical Turk ("mTurk") to provide a standard of comparison for the observational data, and a confirmation of the suspicion that not only 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 (38–40). mTurk qualifications used for this study included: 1) having a HIT Approval Rate greater than 95% and 2) at least 50 Approved HITs. Qualifications are regularly used by mTurk requesters to safeguard against inaccurate and inattentive workers.

2.1.1 Survey instructions

After providing consent, participants were presented the following 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 percentage of how many of these sounds serve a particular function (social or non-social). 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%).

During the survey, participants indicated how often they thought infant vocalizations are 1) directed towards another person (socially directed) and 2) NOT directed towards another person (non-socially directed). Participants answered this question three times with respect to the three ages (3-month-olds, 6-month-olds, and 10-month-olds). Means and standard deviations of these responses were calculated to provide an estimate of general opinions about how often infants use non-socially directed and socially-directed protophones across the three ages.

2.1.2 Survey participants

300 participants completed the online survey, and 239 participants' data were used in final analysis for this study based on correct 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. Detailed demographics of the mTurk survey participants are presented in **Table 1.**

Table 1. mTurk survey participant demographics.

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	**All the time**	19
55-64	24			**Bachelor's**	111	**4+**	12	**Frequently**	46
65+	2			**Master's**	9				
				Doctorate (PhD)	2				
				Professional Degree (JD, MD)	5				

2.2 Part 2: Observational study on the function of infant vocalizations

2.2.1 Data source

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.

To obtain samples of infant vocalizations, we drew from the University of Memphis Infant Vocalization (IVOC) Laboratory's archives of audiovisual recordings. We selected six parent-infant dyads (3 male, 3 female infants) who were previously recorded while engaged in naturalistic interactions and play. 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.

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

All infants completed two recording sessions around ages 3, 6, and 10 months of age.

2.2.2 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 play room with toys and books. In roughly counterbalanced orders across ages, parents were either instructed to interact with the infant (*Interactive* circumstance) or with another adult while the baby was in the room (*Non-Interactive* circumstance). Later at the same age (usually on the same day), the parent was engaged in the other circumstance. 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 (41,42).

2.2.3 Coding for *Interactive* and *Non-Interactive* circumstances

The recordings had been intended to be differentiated neatly as primarily corresponding to *Interactive* or *Non-Interactive* circumstances, but the infants often sought attention from the parents during sessions designated as being *Non-Interactive*, or adults would engage in conversation during sessions intended to be *Interactive*. For this reason, we categorized segments of time within each session as *Interactive* or *Non-Interactive* (often sessions included several *Interactive* or *Non-Interactive* segments of time). These segments were then collated into a single circumstance at each age for each infant to ensure all segments of the recordings were accurately portrayed for analysis of the vocalization data. The amount of time pertaining to each varied substantially, including two segments that included so few utterances (< 5) we did not include them in the analyses (see **Table 3**).

Table 3. Circumstance duration and protophone counts.

Infant	Circumstance	Number of Protophones			Duration (Mins)		
		3 mo	6 mo	10 mo	3 mo	6 mo	10 mo
1	Interactive	446	310	182	53	56	34
	Non-Interactive	4*	47	118	3	7	31
2	Interactive	230	181	108	33	33	33
	Non-Interactive	202	122	70	34	35	20
3	Interactive	311	158	133	35	34	33
	Non-Interactive	163	102	81	36	34	17
4	Interactive	273	103	233	38	17	40
	Non-Interactive	227	384	138	26	42	25
5	Interactive	328	330	89	27	34	34
	Non-Interactive	257	147	117	33	31	32
6	Interactive	442	381	116	60	43	42
	Non-Interactive	13	4*	107	2	1	22
Average		**241.3**	**189.1**	**124.3**	**32**	**31**	**30**

Total duration and counts of the number of protophones for Interactive and Non-Interactive circumstance segments at each age for all infants. Segments marked with an asterisk (*) were excluded from analysis because they included fewer than 5 protophones.

2.2.4 Coding of the sociality of the infant protophones

Coding for circumstance, illocutionary functions, and gaze direction was completed within the Action Analysis Coding and Training software (AACT) (43). This coding software has been used and discussed extensively in previous research from this laboratory (42,44,45). The software affords frame-accurate coordination of video and audio, which is displayed in a special version of the TF32 software (46). 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 had been labeled for vocal type and bounded in time for onsets and offsets in AACT in prior studies (34). 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 was coded within an *Interactive* or *Non-Interactive* circumstance.

All infant protophones that had been previously bounded were also labeled for the present work in terms of illocutionary force (47–49) to indicate potentially communicative functions, which could be easily collapsed into the two **socially directed** and **non-socially-directed** categories. 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 describe vocal acts produced with little or no social intention (34). In this extended usage, vocal play, for example, is treated as an illocutionary force. 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 (34,50). 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.

In our coding of the social or non-social illocutionary functions of infant sounds, we attended to all the contextual information that appeared to be relevant to the judgment of sociality (e.g., gaze direction, gesture, timing with respect to utterances of other speakers, etc.). 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 social acts—otherwise how would humans know when to respond or participate in vocal engagement? If some parents cannot make such judgments, they are surely at a severe 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 their 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.

Non-socially directed protophones were identified as utterances infants produced for their own purposes; such events included vocal play, object-directed sounds, vocal complaints and exultations not directed toward another person, or other protophones produced with no obvious intention or social directivity. Protophones were labeled as **socially directed** 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.

During the coding of sociality, both the primary coder and an independent reliability coder took a broad view of each utterance and its context of production. That is, each time a protophone was located in AACT, the cursors were always stretched so that, during playback before coding for illocutionary force, the coder saw and heard the utterance plus a several-second context both before and after it. If there was ambiguity about how to judge the possible social directivity of the utterance, the boundaries were stretched further until the coder felt confident that no further stretching would improve the coding decision.

2.2.5 Coding for gaze direction of infant protophones

Gaze direction coding was also conducted for all protophones. For this coding, sound was turned off, and the coder determined whether at any time during the vocalization, the infant looked toward another person. The time frame of playback for the protophones was expanded through a special setting in AACT by 50ms before and 50ms 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 (~30ms per frame) and ensured that the entire period of the vocalization was available for visual judgment. For utterances that included no good camera view of the infant (the infant sometimes turned away from the cameras) 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.

2.2.6 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 sequence that has been described in several prior publications (32,34,45). 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 (51); 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 non-socially-directed 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 (34). This completely independent prior coding on half of the data for the present study was available to offer an agreement check on the coding done for the present paper.

3 Results

3.1 Part 1: Opinion survey on the function of infant speech-like vocalizations

Fig. 1 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 non-socially directed. In addition, they thought infants produce fewer non-social 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.

Furthermore, both parents and non-parents reported similar percentages of social and non-social functions. Overall, parents reported infants used social protophones 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.

Fig. 1. mTurk opinion study on social directivity of infant protophones across 3 ages.

[Fig. 1]

Opinions of the survey participants on how often infants use protophones socially and non-socially. Participants believed infants decrease the percentage of non-socially directed protophones between 3-10 months, from 50% at the youngest age to 36% by the oldest age.

3.2 Part 2: Observational study on the function of infant vocalizations

3.2.1 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 circumstances). 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. 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, *Non-Interactive* at 3 months and Infant 6, *Interactive* at 6 months, see **Table 3**). Only 8 protophones occurred in these 2 segments, so the resulting 34 segments provided 6,649 protophones.

To determine if the usage of non-socially directed protophones exceeded that of socially-directed protophones, we used *t*-tests comparing percentages of non-socially-directed protophones against 50% and against the percentages of socially-directed protophones as estimated based on the survey. To test for effects of Age (3 levels) and recording Circumstance (*Interactive* vs. *Non-Interactive*), 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 (52).

Fig. 2 displays the overall percentages of protophones produced by the six infants across the two broad illocutionary groupings of non-socially directed and socially directed. Infants used significantly more non-socially-directed protophones across the three ages than socially-directed protophones, with about 75% of all protophones being non-socially directed. By *t*-tests of the percentage of non-socially directed protophones, it was found they significantly ($p < .001$) exceeded 50% at all three ages and also significantly ($p < .001$) exceeded the percentage of socially-directed protophones estimated by the survey participants at all three ages. **Fig. 2** suggests no notable change in the predominance of the non-socially directed protophones across Age, and indeed the GEE revealed no significant difference in the percentage of protophones that were socially-directed across Age (p = 0.48).

Fig. 2. Social directivity of infant protophones across 3 ages.

[Fig. 2]

Percentage of non-socially directed and socially-directed infant protophones across all observations. Overall, infants primarily produced non-social protophones (75%), suggesting that the great majority of infant sounds are produced endogenously in the first year. Furthermore, a non-significant main effect of Age is consistent with an interpretation of relatively stable use of both social and non-social functions across the three ages compared to the mTurk opinion study results.

Similarly, t-tests of the proportion of non-socially-directed protophones in the two circumstances (*Interactive* vs. *Non-Interactive*, see **Fig. 3**) showed that non-social protophones significantly exceeded 50% in both circumstances (p < .001). Based on the GEE, infants used significantly more non-socially-directed protophones in the *Non-Interactive* circumstance than the *Interactive* circumstance (p < .03), as illustrated in **Fig. 3**. A separate GEE analysis in which only main effects were considered revealed a stronger Circumstance effect (p < .0001).

Fig. 3. Social directivity of infant protophones across two circumstances.

[Fig. 3]

Percentages of social and non-social infant protophones across Interactive *and* Non-Interactive *circumstances. Non-socially-directed 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. For both coders, non-socially-directed protophones predominated, and in fact the reliability coder—who had no knowledge of the hypotheses for this study— showed a slightly higher proportion of non-socially-directed protophones (79.2%) than the primary coder (78.5%).

3.2.2 Protophone usage based on gaze-direction judgments

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

In the earlier study mentioned above (34), 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 the *Interactive* 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 social directivity. In the coding of illocutionary force, momentary gaze direction by the infant toward a person was sometimes not deemed to indicate social directedness. 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 socially-directed 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.

Even though social directedness as determined by gaze-direction did not correspond for as many individual protophones as the illocutionary judgments of social directedness, the overall percentages of non-socially-directed protophones was notably similar for both methods. 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 found the great majority of infant protophones to be non-socially directed. 72% of the infant protophones were deemed not to include person-directed gaze, and 75% were deemed non-socially directed by illocutionary coding.

4 Discussion

Overall, infants used about three times as many non-socially-directed protophones as socially-directed ones. This predominance remained stable across the three ages. Furthermore, even in the *Interactive* circumstance, where parents had been instructed to engage with their infants, non-socially-directed protophones predominated, with twice as many non-socially directed as socially-directed ones. In the *Non-Interactive* circumstance, where parents were engaged in conversation with laboratory staff, the non-socially-directed protophones predominated to a substantially greater extent, with four times as many non-socially directed as socially directed.

The low rate of vocal directivity of the infants in the first 10 months as reported here requires a re-orientation of thinking about the functions of infant protophones. It seems important to draw attention to the fact that all the sessions of recording reported on here were ones where caregiver and infant were in the same room, and where caregivers were aware that they were being recorded. The caregivers also knew that the study was about vocal development, and it was assumed they would endeavor to elicit infant vocalization and thus interaction as much as possible. They also often attended to infant vocalizations even in the designated *Non-Interactive* 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 of the sessions appropriately to the *Interactive* and *Non-Interactive* circumstances. Consequently, we presume parents tried to maximize their infants' socially-directed vocalization— and yet the rate was low.

Partly because the *Non-Interactive* circumstance resulted in a considerably larger predominance of the non-socially-directed protophones, we are suspicious that even more naturalistic recordings might produce an even greater predominance of non-socially-directed 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. The suspicion is supported by recent results where we had the opportunity to compare 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 (53) conducted in the home with the very same infants at approximately the same ages across the first year of life (32). 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 non-socially-directed protophones at home could be considerably higher than we have seen in the present work, since IDS is considerably lower, a possibility that will be explored in subsequent efforts from our group. In future research, we also aim to study a larger sample of infants and to consider more differentiated circumstances of recording.

Our results clearly contradict the apparent standard viewpoint 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. The survey data suggest the general public shares this expectation with the field of child development, assuming babies use protophones for more social purposes than non-social ones.

What is the source of the mistaken impression that non-socially-directed protophones occur far less often than they actually do? 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 interactive vocalizations to a greater extent than non-interactive ones. 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 this particular sound, creating the appearance of, or perhaps initiating social engagements with the infant. Indeed, we have reported evidence suggesting caregivers pay the greatest attention to salient vocal signals such as those occurring in imitation, which is surprisingly rare in the first year (54). Caregivers, and thus people in general, may be inclined to overestimate the proportion of salient vocal signals such as imitation or immediate responses since it seems likely these are the sounds to which parents attend most. So when they render estimates, they tend to overstate the frequency of occurrence of the socially-directed 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 based on randomly sampled segments from all-day recordings at home (32). 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 (55). The 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 (34,50)?

Members of our research group and John L. Locke have argued elsewhere (48,56–58) from an evolutionary-developmental (evo-devo) perspective (59–62) that high rates of exploratory vocalization and vocal play may constitute fitness signals by the human infant. 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—human infant need for such caregiving lasts literally twice as long as in our closest ape relatives (63). 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. Presumably from this point of view, caregivers may then invest more in infants who seem healthy and tend to neglect infants who seem less healthy. Thus, we operate under the assumption that the production of comfortable vocalization can signal well-being and good health. This pattern of fitness signaling may well have applied to the ancient hominin infant, who has been presumed in accord with the hominin "obstetrical dilemma" (64), to have been more altricial than other apes as soon as humans were bipedal. In accord with this reasoning— which proves surprisingly difficult to confirm in the fossil record (65,66)— bipedality had narrowed the human pelvis and required the hominin infant to be born with a smaller head and thus to be more altricial than other apes.

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? 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 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 the 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 the 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 (57). 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 (48).

Play is widely recognized as a theater for practice of the behaviors young mammals will need as they proceed through life (67,68). 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 in the development of language.

5 Acknowledgements

We wish to thank the survey participants, graduate student reliability coders, and the families in Memphis whose infants participated in this research. The research for this manuscript was funded by NIH Grants R01 DC006099, DC011027, and DC015108 from the National Institute on Deafness and Other Communication Disorders and by the Plough Foundation, which supports Oller's Chair of Excellence.

6 References

1. Bigelow AE, Rochat P. Two-month-old infants' sensitivity to social contingency in mother-infant and stranger-infant interaction. Infancy. 2006;9(3):313–25.

2. Bourvis N, Singer M, Saint Georges C, Bodeau N, Chetouani M, Cohen D, et al. Pre-linguistic infants employ complex communicative loops to engage mothers in social exchanges and repair interaction ruptures. R Soc Open Sci. 2018;5.

3. Fonagy P. Affect regulation, mentalization and the development of the self. Routledge; 2018.

4. Liszkowski U. Two sources of meaning in infant communication: Preceding action contexts and act-accompanying characteristics. Philos Trans R Soc Lond B Biol Sci. 2014;369(1651):20130294.

5. Watson JS. Memory and "contingency analysis" in infant learning. Merrill Palmer Q Behav Dev. 1967;13(1):55–76.

6. Winnicott DW. The theory of the parent-infant relationship. Int J Psychoanal. 1960;41:585–95.

7. Oller DK. The emergence of the speech capacity. Mahwah, NJ: Psychology Press; 2000.

8. Goldstein MH, Schwade JA. Social feedback to infants' babbling facilitates rapid phonological learning. Psychol Sci. 2008 May 1;19(5):515–23.

9. Gratier M, Devouche E, Guellai B, Infanti R, Yilmaz E, Parlato-Oliveira E. Early

development of turn-taking in vocal interaction between mothers and infants. Front Psychol. 2015;6(September):1–10.

10. Hsu HC, Fogel A. Infant vocal development in a dynamic mother-infant communication system. Infancy. 2001;2(1):87–109.

11. Bloom K, Esposito A. Social conditioning and its proper control procedures. J Exp Child Psychol. 1975;

12. Routh DK. Conditioning of vocal response differentiation in infants. Dev Psychol. 1969;1(3):219–26.

13. Weisberg P. Social and non-social conditioning of infant vocalization. Child Dev. 1963;34:377–88.

14. Abney DH, Warlaumont AS, Oller DK, Wallot S, Kello CT. Multiple coordination patterns in infant and adult vocalization. Infancy. 2017;22(4):514–39.

15. Gros-Louis J, West MJ, King AP. Maternal responsiveness and the development of directed vocalizing in social interactions. Infancy. 2014;19(4):385–408.

16. Hsu HC, Fogel A. Social regulatory effects of infant nondistress vocalization on maternal behavior. Dev Psychol. 2003;39(6):976.

17. Warlaumont AS, Richards JA, Gilkerson J, Oller DK. A social feedback loop for speech development and its reduction in autism. Psychol Sci. 2014;25:1314–24.

18. Tronick EZ, Als H, Adamson LB, Wise S, Brazelton TB. The infant's response to

entrapment between contradictory messages in face-to-face interaction. J Am Acad Child Psychiatry. 1978;17:1–13.

19. Goldstein MH, King AP, West MJ. Social interaction shapes babbling: testing parallels between birdsong and speech. Proc Natl Acad Sci USA. 2003 Jun;100(13):8030–5.

20. Franklin B, Warlaumont AS, Messinger D, Bene E, Nathani Iyer S, Lee C-C, et al. Effects of parental interaction on infant vocalization rate, variability and vocal type. Lang Learn Dev. 2013;10(3):279–96.

21. Ainsworth MD. Object relations, dependency, and attachment: A theoretical review of the infant-mother relationship. Child Dev. 1969;969–1025.

22. Bowlby J. Attachment and loss. Vol. 1. New York, NY: Basic Books; 1969.

23. Pipp S, Harmon RJ. Attachment as regulation: A commentary. Child Dev. 1987;58(3):648–52.

24. Schore AN. Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. Infant Ment Health J. 2001;22(1–2):7–66.

25. Skinner BF. Verbal behavior. New York, NY: Appleton-Century-Crofts, Inc.; 1957.

26. Watson JB. Psychology as the behaviorist views it. Psychol Rev. 1913;20(2):158–77.

27. Davis K, Panksepp J. The emotional foundations of personality: A neurobiological and evolutionary approach. WV Norton & Company; 2018.

28. Panksepp J. Toward a general psychobiological theory of emotions. Behav Brain Sci.

1982 Sep 4;5(3):407–22.

29. Panksepp J. Toward a cross-species neuroscientific understanding of the affective mind: Do animals have emotional feelings? Vol. 73, American Journal of Primatology. 2011. p. 545–61.

30. Panksepp J, Biven L. The archaeology of mind: Neuroevolutionary origins of human emotions. WV Norton & Company; 2012.

31. Moulin-Frier C, Nguyen SM, Oudeyer PY. Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. Front Psychol. 2014;4:1006.

32. Oller DK, Griebel U, Iyer SN, Jhang Y, Warlaumont AS, Dale R, et al. Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. Front Psychol. 2019;10:729.

33. Moulin-Frier C, Oudeyer PY. The role of intrinsic motivations in learning sensorimotor vocal mappings: A developmental robotics study. In: INTERSPEECH, ISCA. Lyon, France; 2013.

34. Oller DK, Buder EH, Ramsdell HL, Warlaumont AS, Chorna LB, Bakeman R. Functional flexibility of infant vocalization and the emergence of language. Proc Natl Acad Sci. 2013;110(16):6318–23.

35. Harold MP, Barlow SM. Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. Infant Behav Dev. 2013;36(1):84–93.

36. Iyer SN, Denson H, Lazar N, Oller DK. Volubility of the human infant: Effects of parental interaction (or lack of it). Clin Linguist Phonetics. 2016;30(6):470–788.

37. Oller DK, Caskey M, Yoo H, Bene E, Jhang Y, Lee CC, et al. Preterm and full term infant vocalization and the origin of language. Sci Rep. 2019;9(14734).

38. Buhrmester M, Kwang T, Gosling SD. Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspect Psychol Sci. 2011;6(1):3–5.

39. Paolacci G, Chandler J. Inside the Turk: Understanding mechanical Turk as a participant pool. Curr Dir Psychol Sci. 2014 Jun 3;23(3):184–8.

40. Hauser DJ, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav Res Methods. 2016;48(1):400–7.

41. Buder E, Warlaumont AS, Oller DK, Chorna LB. Dynamic indicators of mother-infant prosodic and illocutionary coordination. In: Speech Prosody 2010-Fifth International Conference. 2010. p. 6–9.

42. Warlaumont AS, Oller DK, Dale R, Richards JA, Gilkerson J, Xu D. Vocal interaction dynamics of children with and without autism. In: Proceedings of the Annual Meeting of the Cognitive Science Society. 2010. p. 121–6.

43. Delgado RE, Buder EH, Oller DK. AACT (Action Analysis Coding and Training). Miami, FL: Intelligent Hearing Systems; 2010.

44. Jhang Y, Franklin B, Ramsdell-Hudock HL, Oller DK. Differing roles of the face and voice in early human communication: Roots of language in multimodal expression. Front

Commun. 2017;2(10).

45. Yoo H, Bowman DA, Oller DK. The origin of protoconversation: An examination of caregiver responses to cry and speech-like vocalizations. Front Psychol. 2018;9:1510.

46. Milenkovic P. TF32 [Computer software]. Madison, WI: University of Wisconsin-Madison; 2001.

47. Austin JL. How to do things with words. Oxford, UK: Oxford University Press; 1962.

48. Oller DK, Griebel U, Warlaumont AS. Vocal development as a guide to modeling the evolution of language. Top Cogn Sci. 2016;8(2):382–92.

49. Searle JR. Speech acts: An essay in the philosophy of language. Vol. 626. Cambridge University; 1969.

50. Jhang Y, Oller DK. Emergence of functional flexibility in infant vocalizations of the first 3 months. Front Psychol. 2017;8:300.

51. Oller DK, Wieman LA, Doyle J, Ross C. Infant babbling and speech. J Child Lang. 1976;3:1–11.

52. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Vol. 73, Biometrika. 1986. p. 13–22.

53. Zimmerman FJ, Gilkerson J, Richards JA, Christakis DA, Xu D, Gray S, et al. Teaching by listening: The importance of adult-child conversations to language development. Pediatrics. 2009;124:342–9.

54. Long HL, Oller DK, Bowman DA. Reliability of listener judgments of infant vocal imitation. Front Psychol. 2019 Jun 11;10:1340.

55. Bornstein MH. Infant into conversant: Language and nonlanguage processes in developing early communication. In: Budwig N, Užgiris IČ, Wertsch J V., editors. Communication: An Arena of Development. Santa Barbara, CA: Greenwood Publishing Group; 2000. p. 109–29.

56. Locke JL. Parental selection of vocal behavior. Hum Nat. 2006;17(2):155–68.

57. Locke JL. Evolutionary developmental linguistics: Naturalization of the faculty of language. Lang Sci. 2009;31(1):33–59.

58. Oller DK, Griebel U. Contextual freedom in human infant vocalization and the evolution of language. In: Burgess RL, MacDonald K, editors. Evolutionary Perspectives on Human Development. Thousand Oaks, CA: SAGE Publications; 2004. p. 135–66.

59. Carroll SB. Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom. WV Norton & Company; 2005.

60. Gottlieb G. Developmental-behavioral initiation of evolutionary change. Psychol Rev. 2002;109(2):211.

61. Kirschner M, Gerhart J. The plausibility of life: Resolving Darwin's dilemma. Yale University; 2006.

62. Newman SA, Müller GB. Epigenetic mechanisms of character origination. J Exp Zool. 2000;288(4):304–17.

63. Locke JL, Bogin B. Language and life history: A new perspective on the development and evolution of human language. Behav Brain Sci. 2006;29:259–325.

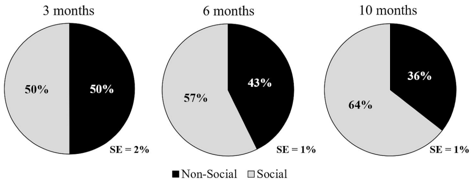
64. Washburn SL. Tools and human evolution. Sci Am. 1960;203(3):62–75.

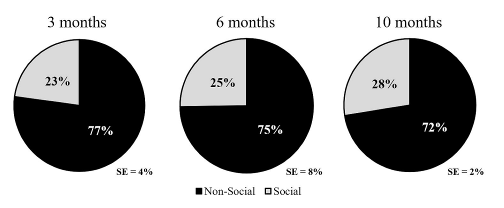
65. Gruss LT, Schmitt D. The evolution of the human pelvis: Changing adaptations to bipedalism, obstetrics and thermoregulation. Philos Trans R Soc B Biol Sci. 2015;370(1663):20140063.

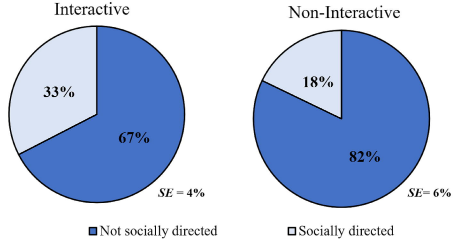
66. Wells JCK, DeSilva JM, Stock JT. The obstetric dilemma: An ancient game of Russian roulette, or a variable dilemma sensitive to ecology? Am J Phys Anthropol. 2012;149:40–71.

67. Bekoff M, Byers J. Animal play: Evolutionary, comparative, and ecological perspectives. Vol. 36. Cambridge University; 1998.

68. Lafreniere P. Evolutionary functions of social play: Life histories, sex differences, and emotion regulation. Am J Play. 2011;3(4):464–88.


3 months
50%
50%
SE = 2%
6 months
57%
43%
SE = 1%
10 months
64%
36%
SE = 1%
Non-Social
Social


3 months
23%
77%
SE = 4%
6 months
25%
75%
SE = 8%
10 months
28%
72%
SE = 2%
Non-Social
Social


Interactive
33%
67%
SE = 4%
Non-Interactive
18%
82%
SE= 6%
Not socially directed
Socially directed