Journal of Research in Health Sciences
ISSN: 2228-7809 | https://jrhsjournal.com

Comparative Analysis of Vocal Patterns in Infant Boys and Girls

Dr. Lukas Schneider[1]*, Dr. Anna Müller[2]
[1]Department of Surgery, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Institute of Medical Research, Ludwig Maximilian University of Munich, Munich, Germany

Main text

Female humans appear to have an advantage in language, from early childhood through late adulthood, reported to include a larger vocabulary, more complex utterances, greater expressive language, and better verbal and pragmatic language comprehension [1]. Wakeful infants produce ‘protophones’ — precursors to speech that include vowel-like sounds, squeals, and growls — at a rate of four or five utterances per minute, more than five times the rate of crying, throughout the first year [2]. The massive number of protophones is in itself surprising, but equally surprising, given the presumed female language advantage, we found that, in the first year, boys produced 24% more protophones than girls. This sex bias was true of infants either at high risk (HR) or low risk (LR) for autism. Both genetic and cultural factors may be involved in this bias, and additional research is clearly called for to investigate the origins of the strong tendency of infants to produce protophones and the unexpected tendency for boys to do so to a greater extent.

Figure 1A shows the highly significant result favoring boys (t-test, $p < 0.0001$) with an effect size (Cohen’s $d = 0.89$) more than four times larger than that typically reported for female language advantage [3]. Both HR and LR boys’ protophone rates were significantly higher than girls’ (HR, $p < 0.005$, boys 27% higher, $d = 1.02$; LR, $p =$

0.01, boys 21% higher rate, $d = 0.78$). Figure 1B displays rates for infants grouped by age, boys showing higher rates at all ages. Figure 1C, D shows results for HR and LR infants, with higher rates in boys at all ages. Generalized Estimating Equations (GEE) tested the Age, Sex, and Risk factors, revealing a Sex effect ($p< 0.0001$) and an Age by Sex interaction ($p < 0.05$), corresponding to a decreasing difference between boys and girls across Correspondence ages (Figure 1B), a pattern driven mostly by the diminishing difference across Age in the HR infants (Figure 1C). Thus, contrary to expectations, protophone rate was higher in boys than girls across the first year, with greatest difference at the earliest ages.

We wondered if the higher protophone rate of the boys would correspond to more rapid development of advanced protophones, namely canonical babbling — baba, mama, and so on — which begins at approximately seven months and involves well-formed syllables that can be used in words [4]. The canonical babbling ratio (CBR) is the number of canonical syllables, such as [ba], divided by the total number of syllables an infant produces, including non-canonical syllables, usually vowel-like sounds. Notably, whereas deaf infants show no reduction in protophone rate, they are sharply delayed in onset and rate of canonical babbling [5]. So protophone rate and canonical babbling may be somewhat independent.

Indeed, boys had no advantage over girls in CBR (Figure S1 in the Supplemental Information), which increased as expected significantly for both sexes across Age ($p <$ 0.005) and Risk (LR higher, $p < 0.05$). Thus canonical babbling, a scaffold for first word acquisition, showed no sex bias, but did show the expected increase with age as well as a higher CBR in LR infants, a finding consistent with prior reports of disruption in canonical babbling of infants with or at risk for autism [6].

We did not set out to study sex effects in speech-precursors, but the longitudinal research reported here afforded us the opportunity to reliably evaluate sex effects through extensive human coding at considerable sample size both of intensive longitudinal home-recordings and of infants. The infants were recorded all day and approximately monthly across the first year (65 boys, $M = 8.55$ all-day recordings; 35 girls, $M = 8.60$) using a miniature audio recorder, yielding ~6800 hours of recording. Twenty-one randomly-sampled five-minute segments

from each recording were coded in real-time by a trained team, yielding >330,000 protophones and >50,000 cries. Coders were blind regarding infant age, sex, and risk status. Coding reliability was high, and discrepancies among coders were small with regard to the effect, indicating boys produced more protophones than girls (see Supplemental Experimental Procedures for methods details and demographics, and Table S1). Cultural factors could contribute to sex differences in protophone rates. But we know of no comparative cross-cultural research on vocal rates of infant boys and girls nor on possible differences in caregiver speech to boys and girls across cultures. A non-significant tendency for caregivers to speak more to boys was seen in our data (see Supplemental Results), and other possible cultural factors could also influence sex differences in infant vocal rates (see Supplemental Discussion).

It is possible that the sex difference is not closely related to language capability — the CBR did not show a sex difference — but rather to a difference in the tendency to vocalize, perhaps owing to sex differences in motoric activity level in infancy [7]. Boys might be said to show higher quantity but not quality in protophone production. Another hypothesis can be formulated in conjunction with a proposed explanation for the high rate (thousands per day throughout the first year) of human protophone production in both sexes (see Supplemental Discussion). The protophones appear to be produced largely endogenously — they are most commonly not directed toward other speakers, occurring at a rate of approximately four per minute even when infants are alone [8]. Even infants born more than two months prematurely and still in neonatal intensive care produce prodigious numbers of protophones [2]. Furthermore, as noted above, there is no sign that deafness reduces protophone rates [5].

This audible endogenous motoric activity, usually produced by infants in comfort, might be motivated by its value as a fitness signal for the altricial human infant, competing for parental investment [9]. One might then suggest that evolution has led to boys signaling their fitness more frequently than girls because they are more vulnerable to death in the first year [10]. This fitness signaling hypothesis could be explored, for example, by correlating parental investment with infant protophone rates. We are, however, seeking other possible explanations for this unexpected sex difference in infant vocal rates (see Supplemental Discussion).

Supplemental Information

Supplemental Information includes one supplemental figure, one supplemental table, Supplemental Experimental Procedures, Supplemental Results, Supplemental Discussion, and Author Contributions and can be found with this article online at https://doi.org/10.1016/j.cub.2020.03.049.

Acknowledgments

This research was funded by grants DC011027 and DC015108 from the National Institute on Deafness and Other Communication Disorders, MH100029 from the National Institute on Mental Health, and by the Plough Foundation, the Holly Lane Foundation, the Marcus Foundation, the Woodruff-Whitehead Foundation, and the Georgia Research Alliance.

Declaration of Interests

The authors declare no competing interests.

References

1. Bornstein, M.H., Hahn, C.-S., and Haynes, O.M. (2004). Specific and general language performance across early childhood: stability and gender considerations. First Language *24*, 267–304.
2. Oller, D.K., Caskey, M., Yoo, H., Bene, E.R., Jhang, Y., Lee, C.-C., Bowman, D.D., Long, H.L., Buder, E.H., and Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. Sci. Rep. *9*, 14734.
3. Hyde, J.S. (2005). The gender similarity hypothesis. Am. Psychol. *60*, 581–592.
4. Oller, D.K. (2000). The Emergence of the Speech Capacity (Mahwah, NJ: Lawrence Erlbaum Associates).
5. Iyer, S.N. and Oller, D.K. (2008). Prelinguistic vocal development in infants with typical hearing and infants with severe-to-profound hearing loss. Volta Rev. *108*, 115–138.
6. Paul, R., Feurst, Y., Ramsay, G., Chawarska, K., and Klin, A. (2010). Out of the mouths of babes: vocal production in infant siblings of children with ASD. J. Child Psychol. Psych. *52*, 588–598.
7. Campbell, D.W., and Eaton, W.O. (1999). Sex differences in the activity level of infants. Infant Child Dev. *8*, 1–17.
8. Oller, D.K., Griebel, U., Iyer, S.N., Jhang, Y., Warlaumont, A.S., Dale, R., and Call, J. (2019). Language origin seen in spontaneous and interactive vocal rate of human and bonobo infants. Front. Psychol. *10*, 729.
9. Locke, J.L. (2006). Parental selection of vocal behavior: crying, cooing, babbling, and the evolution of language. Human Nature *17*, 155–168.
10. Pongou, R. (2013). Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins. Demography *50*, 421–444.

Affiliations and email addresses

D. Kimbrough Oller (koller@memphis.edu)[1,2,3]

Ulrike Griebel (ulrikegriebel@gmail.com)[1,2,3]

Dale D. Bowman (ddbowman@memphis.edu)[1]

Edina Bene (ebene@memphis.edu)[1]

Helen L. Long (hlong@memphis.edu)[1]

Hyunjoo Yoo (hjyoo29@gmail.com)[4]

Gordon Ramsay (gordon.ramsay@emory.edu)[5,6,7]

[1]University of Memphis, Memphis, Tennessee 38152, USA.

[2]Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee 38152, USA.

[3]Konrad Lorenz Institute for Evolution and Cognition Research, 3400 Klosterneuburg, Austria.

[4]University of Alabama, Tuscaloosa, Alabama 35487, USA.

[5]Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, Georgia 30329, USA.

[6]Division of Autism and Related Disabilities, Department of Pediatrics, Emory University. School of Medicine, Atlanta, Georgia 30329, USA.

[7]Center for Translational Social Neuroscience, Emory University, Atlanta, Georgia 30329, USA.

*Correspondence to D. Kimbrough Oller, University of Memphis, 4055 N. Park Loop, Rm 4023, Memphis, TN 38152.

Email: koller@memphis.edu

Figure 1

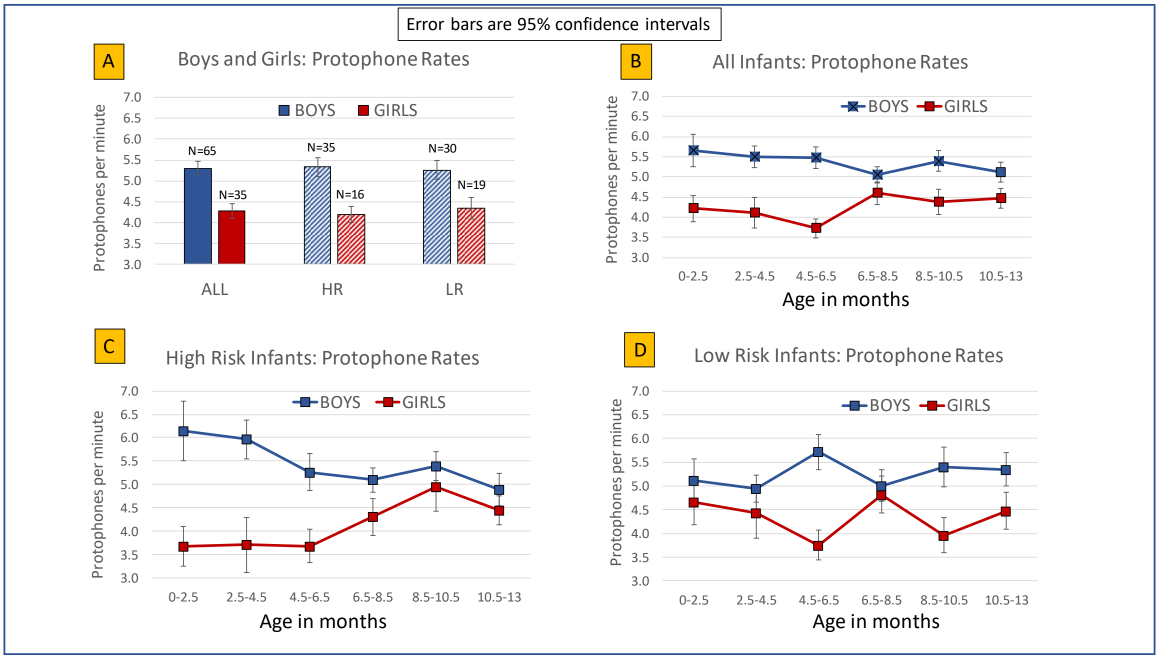

Figure legend

Protophone rates in boys and girls

A. 65 boys produced about one protophone per minute more (~a thousand more protophones/day) than 35 girls (*p*<.0001). The difference favoring boys applied significantly to both infants at high risk (HR) for autism and infants at low risk (LR). Error bars are 95% confidence intervals. Data pertain only to infants who were awake.

B, C, and D. Age analysis revealed that both HR and LR boys produced more protophones at all ages across the first year.

Supplemental Information

Document S1. Supplemental Experimental Procedures and One Table, Supplemental Results and One Figure, and Supplemental Discussion

Supplemental Information

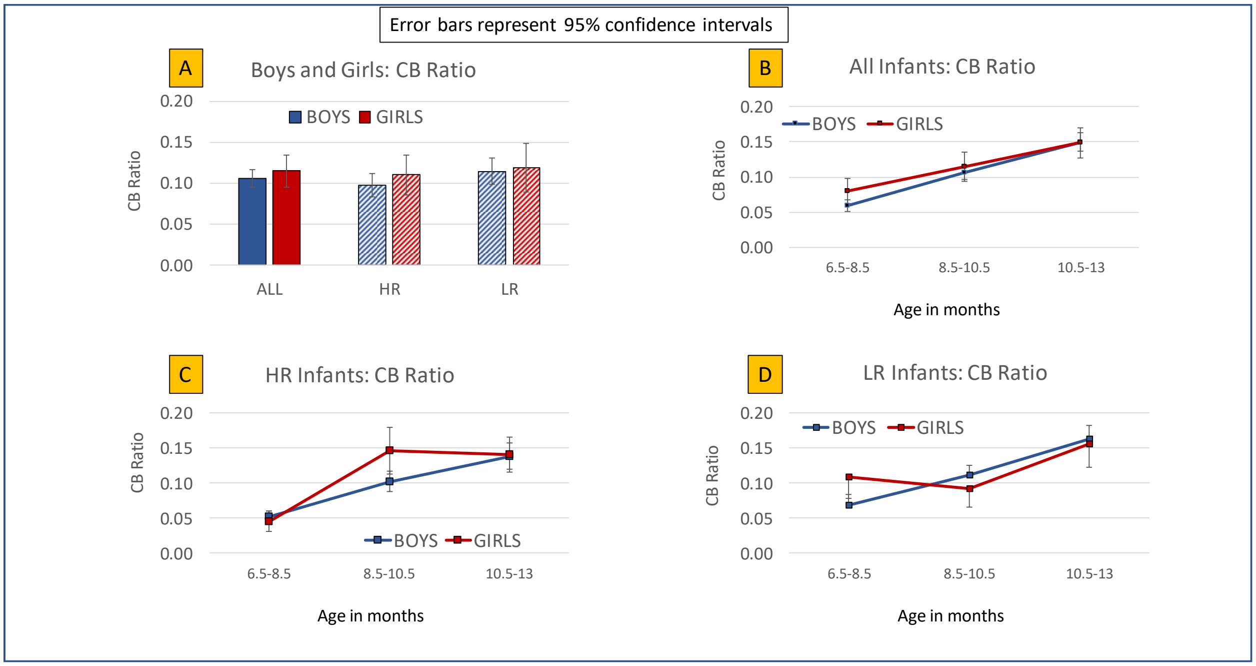

Figure S1: Canonical babbling ratio (CBR) in boys and girls (related to explanatory material in the main text about canonical babbling and to **Canonical babbling ratios as reflected in the coding for boys and girls**, below under Supplemental Results)

Figure S1 Legend

A. In contrast with protophone rate (i.e., volubility of all speech-like utterances), boys did not show any advantage in canonical babbling ratio (CBR), which is a measure of the *degree* to which the protophones of an individual infant are speech-like. CBR can be thought of as providing an indication of the level of vocal development of infants beyond 6 months of age, since utterances that include canonical syllables are the most advanced kinds of protophones, providing the raw material for words. Figure S1 does not display data prior to six months because all values are 0 or nearly 0 below that age, given that canonical syllables rarely occur before then. The error bars represent 95% confidence intervals.

B, C, and D. Analysis by GEE [23] revealed that, of course, CBR increased with Age, because infants acquire increasing capability for producing more speech-like utterances as the end of the first year approaches, the point at which first words typically appear. But boys and girls showed similar CBRs, with no significant differences.

Table S1: Demographics and key data for the infant sample related to methods of the main text and **Participants in the all-day recordings**, in **Supplemental Experimental Procedures**

	LR		**HR**	
Sex	**M**	**F**	**M**	**F**
N	30	19	35	16
Prot/min, Mean	5.26	4.35	5.33	4.20
SD	1.20	1.14	1.38	0.76
95% CI	0.44	0.53	0.46	0.38
Cry/min	0.84	0.54	0.80	0.85
SD	0.54	0.35	0.49	0.71
95% CI	0.20	0.13	0.16	0.24
CBRatio^	0.11	0.12	0.10	0.11
SD	0.08	0.13	0.07	0.09
95% CI	0.03	0.06	0.03	0.05
IDS/min	2.50	2.35	2.70	2.04
SD	1.52	1	1.57	0.80
95% CI	0.55	0.37	0.53	0.27
SleepSegs	27%	25%	25%	32%
SES Low	44%	44%	50%	56%
BirthRank*	2.17	2.47	2.51	2.5
SD	0.65	1.22	0.66	0.63
BrthHght(cm)*	20.5	19.9	20.3	19.9
SD	0.80	2.54	0.90	0.82
BrthWt (lb)*	8.31	7.81	7.76	7.16
SD	1.07	1.20	1.04	1.02
BrthHdCirc (cm)*	35.6	33.5	34.3	34.5
SD	1.32	1.04	1.41	1.82
Race:White**	87%	84%	57%	44%
Race:Black**	7%	11%	20%	31%
Race:Asian**	0%	0%	6%	0%
Race:Other**	7%	5%	17%	25%
Ethnicity:Hispanic**	7%	16%	11%	6%
Ethnicity:NotHisp**	93%	84%	89%	94%

Table S1 Legend

Prot/min = Protophones per minute in segments where infants were awake averaged across ages **Cry/min** = Cries per minute in segments where infants were awake

CBRatio = Canonical babbling ratio (# of canonical syllables / # of all protophone syllables) in segments in infants who were awake

IDS/min = Number of utterances of infant directed speech per minute in segments where infants were awake

^CBRatio was computed as the mean for the 3 recording ages pertaining to the second half year only (6.5-8.5, 8.5-10.5, and 10.5-13 mo), because prior to that time, CBRatio is typically 0.

*Values are those available for analysis as reported by parents through parent questionnaires administered at enrollment. Missing data occur where data were not fully available (e.g., head circumference measurements), or where parents chose not to self-identify (e.g., race other than white/black/Asian).

The **Race and **Ethnicity** values are based on the categories required for reporting to the National Institutes of Health. Race and Ethnicity are treated as fully independent, with the only options under Ethnicity being Hispanic/Latino and non-Hispanic/Latino. Some parents chose not to self-identify for Race, and were thus categorized as Other, which also included self-identifications of mixed Race.

Supplemental Experimental Procedures

SEP1: Approvals

This research was conducted at the University of Memphis and at the Marcus Autism Center. The protocol was approved by Institutional Review Boards of the University of Memphis and Emory University. All methods were performed in accordance with the guidelines and regulations of the IRBs. Signed informed consent was given by parents of all the infants.

SEP2: Participants in the all-day recordings

Recruitment: The 100 infants were recruited in a large-scale longitudinal study on development of social engagement over the first three years. HR participants with an older sibling diagnosed with autism and LR controls with no family history of autism were recruited at birth. All these infants were recruited by advertising and/or word-of-mouth, at birth or during pregnancy, and were enrolled after consent being granted by parents in accord with the Emory University approved protocol. The infants came from families who were of relatively high socio-economic status (SES) as determined primarily by mother's educational level—a median split was used to provide data in Table S1 below. Perhaps because the SES factor was limited in range across the infants, analysis by GEE showed no important SES effects. Consequently, the SES factor was excluded from subsequent analyses.

Birth order, another potentially important factor in vocal development and interaction, showed very little variation among the infants' families. Because the investigation is an infant sibling study, infants in the HR group are by definition, later borns, and the LR infants were selected to roughly match them, as seen in Table S1.

The longitudinal study involves many components, but for the present purposes, the relevant part is the all-day recordings that were scheduled to occur for all the infants, starting optimally in the first month after birth. The study of these was planned to include all-day recordings at every month starting at 0 months; an average of 8 (range 5-12) recordings were obtained from the infants over the first 12 months, with an average recording time of ~11 hours per day.

Demographics for the sample along with data for coded parameters are presented in Table S1, above (and under Supplemental Results below, Cry/min, CBRatio, and IDS/min are discussed in more detail; all human coded data pertain to segments where infants were deemed awake by the coders, see **Coding** below). The SDs reported for the coded parameters in Table S1 reflect the fact that all the parameters involved overlap within both the HR and LR groups between boys' and girls' values.

Why there were more boys than girls in the sample: The sex imbalance (65 boys, 35 girls) was unexpected; it appeared to be the product of self-selection by the individual families recruited in the Emory University/Marcus Autism Center study and has been observed for more than 8 years, with no current explanation. The sampling imbalance favoring boys occurred in both groups (High Risk, 69% boys; Low Risk, 61% boys). One possible explanation for the imbalance is that parents of boys tended to be more concerned about their infants' health and welfare than parents of girls. Boys are indeed more vulnerable to death, especially in the first year according to prior literature [1-3]; thus perhaps parents more often volunteer for longitudinal investigations where boys' health and welfare will be evaluated repeatedly. Still, we know of no such sex imbalance having been reported in other longitudinal research projects.

SEP3: Device and recording procedure for the all-day audio recordings

The battery-powered LENA recorders [4] are about the size of an iPod, and for full-term infants, fit in the chest pocket of special clothing so that mouth-to-microphone distance is nominally 5-10 cm. The system can record up to 16-hours. The sampling rate is 16 kHz, yielding recordings that are quite satisfactory for the human coding and analyses presented in the present article. Automated analysis of the LENA recordings is obtainable, but the present work relies instead on human coding, which is more reliable although more time consuming by far.

The recording procedure was standard for the LENA system. Briefly, parents were provided with the appropriate clothing, were given instructions regarding how to turn it on and place it in the child's clothing at wake up time, then how to turn it off at bedtime and return it to the staff of the research project for processing.

SEP4: Recordings and dataset

After a single all-day recording was obtained, we extracted the waveforms and randomly selected (RS) 21 five-minute segments for human coding and analysis from each recording, avoiding the first and last segments of the day in each case. Altogether there were 859 all-day recordings, from which (21x859) 18,039 five-minute segments were human coded. To afford the possibility of optimally comparing vocalization rates across age, we took a measure of sleep through a questionnaire responded to by coders at the end of coding each 5-minute segment (Was the infant asleep? Scale 1-5). 387,262 infant utterances (protophones+cries) were identified by the coders in these RS segments as pertaining to awake infants. Our analyses as presented in the main text pertain to these awake segments.

SEP5: Coding categories

*Phase 1 coding***:** Phase 1 coding focused on the presumed precursors to speech, the protophones [5], but also provided counts for cries, which were treated as the other most prominent kind of sound produced by infants in the first year. Laughs were coded as well, but they were quite infrequent, accounting for <1% of all coded utterances. Other coded utterances (whispers, ingresses, other sounds [which included clicks, isolated fricatives, bursts, and so on], also accounted for an extremely small proportion of the utterances). Consequently, our analyses focused on protophones (85% of all utterances) and cries (13% if all utterances) only. The three precanonical protophone types that were coded (squeals, growls and vowel-like sounds) were collapsed together to simplify the analyses.

The reason the distinction between cry and protophones is crucial for this work is that cry maintains an obligatory strong connection with the expression of distress across early infancy, and includes both acoustic and facial characteristics marking it as cry. Protophones, on the other hand, do not show an obligatory connection with any particular affect condition at any age. In fact all the protophones we have examined (squeals, growls, vowel-like sounds, raspberries, voluntary ingressive sounds, frication sounds, and both marginal and canonical babbles) have been reported by parents and/or shown in our coding data to be used on different occasions with positive, neutral, *and* negative affect [6, 7].

The protophones are spontaneous in the sense that no particular emotional condition or external stimulus is needed to inspire their production, and in this way they appear to provide the groundwork for the development of the speech capacity—all elements of speech must be free of any particular emotional content, in that speakers must be able to produce them in any emotional state.

A recent study from our laboratory [8] addressed the possibility that an analogous distinction between cry and protophones might exist in one of our closest phylogenetic relatives, the bonobo, *Pan paniscus*. Protophone-like sounds did seem to occur in the three captive bonobo infants (always with their mothers) that were studied, although they were used with more restricted functions and at far lower rates than in the human infants. Infant bonobo screams were deemed analogous to human crying, although we observed no sounds that were similar to prototypical human crying from an acoustic perspective. Other research suggests very low rates of protophone-like sounds in chimpanzees [9, 10]. The massive rates of protophone production in human infants by comparison with rates of similar sounds in our closest phylogenetic relatives suggests, consequently, that human infants may have long been under selection pressure on both the capacity and the inclination to produce protophones, and thus also, on differentiating protophones from cry. As a result, maintaining counts of both in our research is considered critical.

Coders were encouraged to use their human endowment for recognizing speech and speech-like sounds when they differentiated cry from protophones (the task involved a forced choice), since we operate on the assumption that human listeners have a natural capacity for drawing the distinction [11]. Indeed, as will be seen below, coder agreement on the distinction is good, although some utterances mix features of cry and protophones, yielding intermediate cases. Yet if it were *not* possible regularly to draw the distinction between cry and non-cry, caregivers would be handicapped in assessing infant needs—it seems clear that selection pressure throughout hominin history must have been applied to making cry a signal of distress and to leaving the protophones free to vary in emotional expression. But simply to define cry as a distress signal is insufficient to differentiate it from any of the protophone types, because these have all been shown to manifest positive, neutral, *and* negative *facial* affect on different occasions of usage, all three by not later than 3 months [6], when positive affect is first systematically manifest in smiling. Furthermore, negative *vocal* affect is also discernible at good reliability in all the protophones on occasions of usage where it can be determined from facial affect [12].

Both cries and protophones (see [8, 13] for details on how the distinction between cries and protophones is made) are counted within this system in accord with a "breath group" criterion [14], where each voiced period produced on a single egress is counted as one utterance. Also each ingress or each cessation of voicing that is long enough for an ingress (whether one is heard or not) is deemed an utterance boundary, although a glottal hold followed by further voicing without ingress, either in cry or protophones, no matter how long, does not terminate an utterance. Cry and protophone utterances are thus counted in an analogous way in accord with the breath-group criterion, which breaks cry bouts into individual cries of roughly similar dimensions to protophones produced in bouts.

Phase 1 coding Questionnaire: After the 5-minute listening and coding period for each segment, the coders in Phase 1 responded to a number of questions; the key ones were as follows. 1) Did any other person talk to the baby? This could be the parent or another adult or child. 2) Did any other person talk to someone else? For example, two adults might have talked to each other, or a person might have talked on the phone. 3) Do you think the baby was alone in the room? And 4) do you think the baby was asleep? The questions were answered on a 5-point scale, where 1 indicated never, 2 some of the time, 3 about half the time, 4 most of the time, and 5 the entire time.

Phase 2 coding: After coding the 21 segments of a recording in Phase 1, the coders were automatically assigned (by software incorporated into the AACT system, see below) to code eight of those same segments again. The eight were selected on the basis of the protophone count from Phase 1 coding as well as the degree to which there had been Infant-Directed Speech (IDS) as reflected in the questionnaire answers to item 1 above. The coding of these 8 was also conducted in real-time. There were two tasks in Phase 2 coding: 1) to count both canonical and non-canonical syllables produced by the infant wearing the recorder, and 2) to count utterances of other speakers, labeling them as either IDS or ADS (speech directed to someone other than the infant wearing the recorder, thought of usually as adult-directed speech, but sometimes it was speech directed to another infant or child).

The definition of canonical syllables is well-established in the literature on infant vocal development [15], having been applied in a wide variety of studies worldwide [16-18]. The definition is primarily intuitive as with the distinction between protophones and cries, and for the same reasons. The definition makes reference to "well-formed" syllables that could be used in real words, syllables such as [ba], [ma], or [da]. The onset of canonical babbling is typically recognized most prominently when infants produce repetitive sequences of canonical syllables, such as [bababa] or [dada]. However, it is important to note that even after the onset of canonical babbling, non-canonical syllables still predominate among the protophones for months. Even at 12 months, a typical canonical babbling ratio is on the order of .1 to .2 (see comments below in Supplemental Results and Figure S1).

Even at relatively low CB ratios, parents must be able to recognize canonical babbling when it begins, typically at 6-7 months, because if they could not, they would be at a notable disadvantage in guiding their infants' word learning. It is well known that when infants begin to produce canonical babbling, parents intuitively begin suggesting meanings for the infants' utterances [19]. The infant who says [baba] may be guided to use [baba] to label something that way, such as "bubble" or "bottle". High coder reliability has been reported in studies of canonical babbling [20, 21], indicating that counting of canonical and non-canonical syllables is very trainable.

Exclusion of sleep segments in calculations of per minute rates of all vocal types: All the calculations related to amount of vocalization (including canonical babbling) and amount of IDS or ADS were made in such a way that segments where infants were deemed asleep were excluded (see Table S1 for average values for each of the four infant groups). Sleep is more frequent in younger infants: 31% of segments at 0-2.5 mo were deemed to pertain to an infant asleep, while 23% were deemed asleep at 10.5-13 mo. Breaking the data down by sex, at 0-4.5 mo boys showed 30% and girls 31% sleep segments, while at 8.5-13 mo boys showed 23% and girls 24% sleep segments. The inclusion of only awake segments in the calculations of vocalization rates supplies a normalization across age; all data in our figures are based on amounts or proportions of events occurring during wakeful periods.

SEP6: Coding procedure

The decision to use human coding was made on the grounds that no automated system can yet approach the capabilities of human listeners in judging speech and speech-like sounds in naturalistic circumstances involving the noises and multiple voices of the home. The LENA system's automated analysis, which is state of the art, is modeled entirely on human-coded samples. Still, its reliability is judged in terms of how well it approximates the human-coded values, and the first author (an original and continuing member of the Scientific Advisory Board of the LENA Foundation) can attest to the fact that everyone associated with development of the system acknowledges human coding as the gold standard. The goal of the present work is to obtain maximally reliable estimates of infant volubility. Consequently we used human coders for the work.

In order to make it possible to acquire a large human-coded dataset, we implemented a real-time coding method. The data collection at all ages of the present study involves listening to each assigned 5-minute segment in real-time and pressing the appropriate key in the AACT (Action Analysis, Coding and Training) [22] computer-based coding environment for each of the infant protophones (vocants, squeals, growls) plus each occurrence of cries or laughs. An "other" category plus categories "whisper" and "ingress" were also allowed, but like laugh, they were rarely used, and were not included in the analyses below or those reported in the main text.

SEP7: Coders and training

Composition of the team: The coding team consisted of 16 normally-hearing female Masters students in Speech-Language-Pathology from the University of Memphis School of Communication Sciences and

Disorders. The students in the program are overwhelmingly female, and they could be hired as coding team members for a two-to three-year period at a cost that was affordable to the grant and other available resources. We have, over the years, had a few male coders. There has never been any indication that the male coders responded to the coding tasks in an unusual way (coding agreement has been similar to that of the female coders), although a direct comparison of possible different reactions to infant sex has not been possible, because *all* the coders in the present study were female, and it is the first one with a large enough sample size of infants to have usefully examined sex differences in infant vocalization. The coders were not informed that there would be any analysis by sex, and they were not provided with any information about the infants other than the recordings themselves.

Training: All the coders had been trained in phonetic transcription during their program of study. Training for the coding of infant vocalizations began with a two-hour introduction to infant vocalizations with both audio and audio-video examples presented primarily by the first author, whose work for decades has been focused on infant vocal development and the establishment of appropriate descriptive schemes for prelinguistic sounds.

The coding team was also introduced in this initial session to AACT. The system, under development since the 1980's, affords numerous conveniences for coding in both real-time and repeat-listening, and participants learned what they needed to learn to use the AACT system well enough to begin coding their training samples within an hour or so. Thereafter, a series of practice tests were given during which each coder worked independently or with a senior coding partner (a PhD student or another member of the coding team who had gone through the training in a previous year), coding infant vocal samples much like the ones that would be coded when the real data collection began. Several such rounds of coding and reviews of the results with the training staff occurred during a period of 6 to 8 weeks. These reviews consisted of at least one meeting per week of the group as a whole and individual meetings occasionally of the first, fourth, fifth, or sixth authors with individual coders whose results seemed to warrant special attention and/or correction.

In order for a coder to be deemed ready to begin coding for data collection in the study, it was required that their final rounds of coding fall within 10% of the protophone and cry counts in a gold standard set of five sessions where the original coding had been done by author one. Also the coders were required to have diverged by an average of not more than 1 on the scale from one to five on the questionnaire items pertaining to IDS, ADS, Alone, and Sleep. A check on the agreement levels for the coders was made at the end of this period of training for each (to be reported below) based on 88 five-minute segments from 8 different infants that were not among the 100 infants in the present study.

SEP8: Coding protocol and organization

Assignments: Each infant was assigned to a specific coder; coders were assigned to multiple infants, in clusters of 4 infants. 9 of the coders received two clusters of 4 to code in different years (the coding was done across a two-year period), while 7 were assigned to a single cluster of 4 infants. This yoking of coders to infants applied without exception for 92 of the 100 infants. For the remaining 8 infants, the recordings were split between two coders at random—the split was necessitated by the loss of two coders in midstream of the coding enterprise, one in the first year of coding and one in the second. For the infants that had been assigned to those two coders, a few of their recordings had to be reassigned to other coders. As a result, each of the 859 all-day *recordings* remained yoked to a single coder.

The protocol specified that coders should work through the entire data set for each infant to which they had been assigned before proceeding with the next infant. Furthermore coding of each recording was completed before coding of another recording was begun, and the 21 five-minute segments were coded and questionnaires were answered for each segment in Phase 1 coding according to the chronological order in which they had occurred during the recording day. After coding Phase 1, Phase 2 assignments were automatically generated by AACT, and the coders continued with the 8 segments that had been selected, coding for canonical babbling and IDS, again with each segment being coded according to its chronological order during the recording day.

Blinding: The order of coding for recordings within infant was randomized to prevent the coders from knowing the age of the infants directly. Furthermore the coders were blind to the fact that they had been assigned to HR (high risk) or LR (low risk) infants in roughly equal amounts, and the distribution within coder assignments of sexes of infants was also random and not revealed to the coders. Thus coders were blind with respect to Age, Risk, and Sex except to the extent that they could guess these conditions based on their listening to the segments during the coding task.

Justification for yoking of coders to infants: The reason for yoking of coders to infants is that one learns, while listening, to recognize speakers, family patterns, and vocalization patterns within infant. This experience helps make the coder more confident and reduces the stress of the coding, a task that requires considerable concentration. We presume coding is more accurate when coders are yoked to infants. Furthermore, coder biases complicate within-infant interpretation (for example for changes across age) if there are multiple coders for a single recording or infant; thus all 21 five-minute segments from an individual infant recording were coded by the same individual, and all recordings available for that infant were coded by the same individual as indicated above, except for the 8 infants whose coders left the team in midstream. The five-minute coding period for each segment was followed by about one minute for the questionnaire that followed and for saving of files.

SEP9: Coder agreement

Phase 1 coding agreement: The primary issue at stake in assessing coder agreement is its relation to the key findings of the study, in particular the finding that protophones were produced more by boys than by girls. To assess the likelihood that individual coder characteristics (e.g., one coder tending to code more utterances overall than another coder or tending to code more utterances as protophones rather than cries than another coder) might have exaggerated the findings or might have yielded an unreliable outcome (i.e., a new set of coders would not have likely produced data replicating the pattern of results), we determined the agreement levels on numbers of protophones and cries coded across pairs of coders who independently coded an identical set of segments. The agreement statistics of most interest concern the degree of discrepancy between the results of any pair of coders on the number of protophones (and to a lesser extent of cries) determined through the coding of each, as well as the correlation of segment by segment results for the two coders.

All coders coded (almost all of) 88 five-minute segments independently. These represented randomly-selected segments from eight all-day recordings, all from different infants, none of whom were among the infants in the study. These 88 segments were part of a post-training set designed to document agreement, that all coders working in our laboratories for the past four years were required to complete as a final step in the training process. These 5-minute segments were drawn from the larger database of recordings obtained on HR and LR infants in the Atlanta Marcus Autism Center project.

We made a comparison on this set involving 7 coders paired with 7 others, resulting in strong evidence that coding biases or other effects would have been very unlikely to have played any significant role in the main finding of the study (namely the results showing boys produced more protophones than girls). The correlations between the average numbers of vocalizations coded by the 7 coders in the first group with the 7 coders in the second was r=.97 and .99 for protophones and cries respectively. Mean correlations for vocalization counts for individual pairings of the coders *across* the two groups (n ~88 for each of 49 pairings of coders) were r=.86 for protophones and .85 for cries.

More important, both the 7 individuals from the first and the 7 individuals from the second coding group found large numbers of protophones (means of 3.24/min and 2.97/min) in the ~88 segments that were coded by both groups independently. These values are lower than the estimates in the Figure of the main text, because the agreement sample included a few randomly-selected sleep segments, while the data in the main text excluded them. To place these values in the perspective of the effect sizes for boys vs. girls on protophones, notice that the difference in average numbers of protophones per minute found by the two groups of coders was about 9%, but the difference between boys and girls was 24%, more than two and a half times larger.

Both groups of coders that were tested for across-group agreement were also subjected to within-group comparisons, where every possible pairing of 7 within each group (n = 21 in both cases) was tested. The results for agreement were also supportive of the idea that coder differences could not have invalidated the main study findings. The correlations between numbers of protophones and numbers of cries across the 21 pairings of coders from the first coding group was .86 and .81 (n~88 for each of the 21 pairings) respectively; for the second coding group, the protophone and cry correlations were .83 and .81.

A final agreement analysis was conducted by randomly assigning 7 coders to recode segments that had previously been coded by a primary coder (namely another member of the same coding group) in a separate research project involving recordings of infants not involved in the present research, all of them typically developing infants, all recordings made in Memphis. All the coders in the group received such agreement coding assignments, which were semi-randomly made, ensuring that every agreement coder was paired with every other coder for some segments and ensuring that segments from an individual infant's recordings would be coded by the agreement coder in clusters, with segments occurring in the same order in which they had occurred for the primary coders. Also, both primary and agreement coders finished an infant's data before going on to another infant's data, although they were both blinded about how the assignments had been made. Also every coder received assignments for recordings from a variety of ages. 212 segments were assigned for this agreement study. The results showed a .89 correlation on protophone counts between the first round of coding and the second (by two different coders for each segment). The overall picture was very much as in the other agreement assessments: for example, both the primary coders and the agreement coders showed very large discrepancies between the number of protophones and the number of cries, with the primary coders indicating 8 times more protophones than cries in the segments assigned, and the agreement coders indicating 10 times more protophones than cries.

More importantly for the present study, on both the first and second coding of each segment, the coders showed very large numbers of protophones, 3.5-3.6 protophones per minute, a value somewhat lower than values in the main text, presumably again because the agreement segments were selected semi-randomly and thus included some segments where infants were asleep. Notably the difference between the two passes of coding was even smaller than in the group evaluation described above. The two passes yielded differences of only 3% between the numbers of protophones counted in the two cases, a value 8 times smaller than the difference between protophone rates found for boys and girls (24%) as reported in the main text.

This 212-segment agreement study was also used to assess the questionnaire results, which were based on scaled judgments ranging from 1 to 5. The questionnaire item relevant to the present study concerned sleep. The correlation across the primary coders and the agreement coders was .76 for the judgments about an infant being asleep.

Phase 2 agreement study: Using the 212-segment study for the data on agreement in Phase 2 coding, we found that the coders on the two separate and independent passes for each segment correlated with each other at .89 for CB Ratio. The CB Ratio values differed across the two occasions of coding by only 3% (mean CB Ratio across the 212 segments of 0.0411 for the first coding and 0.0422 for the second).

SEP10: Statistical treatment

The analyses regarding all the variables in the Figure of the main text were conducted with unpaired two-tailed t-tests as an initial assessment of boy (N=65) vs girl (N=35) differences. Boys exceeded girls on protophone rates by this measure at $p<.0001$ but did not differ significantly from girls on CB Ratio (see Figure S1).

However, the design of the research allowed a more elaborate analysis accounting for not only Sex but also for Risk level and Age. These analyses were conducted with Generalized Estimating Equations (GEE) [23] implemented in R. GEE is an advanced form of modeling where fixed and random effects can be accounted for, but where there are additional advantages over traditional mixed models approaches. GEE is preferable over traditional mixed models frameworks for semi-longitudinal research when there are correlations among data from participants across conditions, and when the number of observations varies for participants within or across conditions (for example, the number of recordings per infant varied from 5 to 12, and the number of segments available within recordings varied substantially because among the 21 that were selected for coding at random, those where the infant was deemed asleep by the coder were eliminated from the GEE analysis). In essence the approach offers an assessment that estimates, on a principled basis, the means and standard deviations relevant for the analysis while taking into account intragroup correlations and variations in numbers of observations. The GEE approach also has the advantage of requiring no normality assumption.

The GEE results supported the main conclusion associated with the preliminary t-tests, importantly also revealing that boys produced more protophones than girls at $p<.0001$. In addition, however, the GEE analysis suggested an interaction discussed in the main text and again below, indicating that the difference between boys and girls in protophone rates diminished across ages.

Supplemental Results

Canonical babbling ratios as reflected in the coding for boys and girls

As indicated in the main text, we tested the possibility that the higher protophone rate of the boys in our sample would also be reflected in more rapid development of more advanced protophones, namely canonical babbling (e.g., baba, mama…), beginning ~7 months and involving well-formed syllables that can be used in words. The canonical babbling ratio (CBR) is the number of canonical syllables, such as [ba], divided by the total number of syllables an infant produces, including non-canonical syllables, usually vowel-like sounds. While the distinction between canonical babbling and precanonical babbling (both are termed "protophones" in the terminology of Oller [5]) is clearly important, it is also important to emphasize that both types of protophones are foundations for speech. Neither constitutes speech, but both supply vocal features (flexible phonation and phonatory learning in the earlier protophones, well-formed syllabicity and syllable learning in the canonical protophones) that are required in speech, and further these features appear to be absent in non-human apes and for the most part in primates in general— interestingly certain new world monkeys may be the most vocally capable primates other than humans [24-28].

One might hypothesize boys as more vocal in *phonatory activity*, but girls as more vocal in the *articulatory realms reflected in canonical babbling*. But Figure S1 (above) shows boys had no advantage over girls in CBR. Figures S1A-D reflect significant effects revealed by GEE of Age (increasing CBR with Age, $p<.005$) and Risk (LR higher, $p<.05$) and a significant Age by Risk interaction (the youngest and middle Age groups were differently related to Risk, $p<.01$). These significant differences are predictable based on prior literature, but there has never been any prior test to our knowledge of CBR across the sexes. In our data, canonical babbling, a key scaffold for first word acquisition [5], showed no sex bias in the sAllecond half year of life.

Data on cry rates in the sample

As suggested previously, boys' higher protophone rates might have been to be due to higher general activity level in boys [29] (along with other possibilities), in which case crying might also be predicted to occur more often in boys. Cries were counted with the same breath-group principle that was used in counting protophones.

As expected based on recent empirical research, cry rates were far lower than protophone rates, by a factor of ~5, and declined notably across the year. There was a 20% higher rate of crying in boys ($d=0.26$), but it was *not* statistically significantly. The data suggest the non-significant sex difference occurred primarily in the LR infants and was large only at the youngest age (0-2.5 mo). GEE indicated a highly significant effect of Age ($p<.0001$), but not of Sex. However, there were significant interactions: Sex by Risk ($p<.03$), reflecting the fact that HR girls cried only slightly more than boys ($d=-.07$), while LR boys cried considerably more than girls ($d=0.65$), and Sex by Risk by Age ($p<.02$). Thus crying, deemed unrelated to language, showed no overall sex bias in our data, but suggested possible genetic and developmental bases for sex and autism risk interactions. Further assessment of these effects will likely require additional research at larger sample sizes.

Infant-directed speech (IDS) and Adult-directed speech (ADS) in the sample

Infant-directed speech (IDS) is thought to be a key factor in early speech development [30-33]. Whether parents preferentially interact communicatively with girls or boys (infant-directed speech, IDS) is debatable based on existing literature [34-37]. GEE on the data from the present study indicated a statistically non-significant difference ($p<.07$) favoring talk to boys (IDS). There were no differences even approaching statistical significance in ADS (where both adult talk and talk of other speakers were taken into account) in the presence of boys and girls. Thus, we found no persuasive evidence that the sex biases were attributable to social interaction or environmental talk differences, but the sample size may not have been large enough to have detected a true meaningful difference. Further research is of course highly desirable.

Additional post hoc analyses on protophone rates of boys and girls in the first and second halves of the first year

To add perspective on the Age by Sex interaction ($p<.05$) for protophone rates as reported in the main text, we conducted additional analysis. We pooled the data on protophone rates at 0-6.5 mo (the first 6 mo) for all infants, segregated for HR only and for LR only, and similarly for 6.5-13 mo data (the last 6 mo), by using the simple method of calculating means for the three component age groups in each case. We found that for all infants (both HR and LR), the boys showed about 38% higher rates than girls in the first 6 mo, but only 16% higher in the last 6 mo. For HR infants, boys were 57% higher in the first 6 mo, and 12% higher in the last 6 mo. For LR infants, boys were 23% higher in the first 6 mo, and 10% higher in the last 6 mo.

A positive correlation between the protophone rates for the first six months and the second six months would suggest consistency within infants regarding protophone rates across time, and the magnitude of the correlation would offer a measure of the effect size of that consistency. For all the infants (boys and girls, HR and LR), the correlation for the protophone rates was $r=.40$, $p<.00001$, a highly significant though not a very high correlation. For boys, $r=.35$, $p<.005$; for girls, $r=.32$, $p<.07$. The results suggest a moderate degree of within-infant consistency in protophone rates across time.

Supplemental Discussion

Prior research on volubility by sex in humans and non-human primates

Study of volubility in non-human primates has been relatively rare (see below), but there does exist one direct comparison across volubility of male and female infants in the East African vervet monkey (*Cercopithecus aethiops pygerythrus*) [38]. Notably females were more voluble for both of the two categories of sounds that were monitored. The pattern seen in the vervet monkey parallels the prior reports on humans cited by the authors, where it has been generally found that females are more voluble than males. Quoting from Locke and Hauser [38]: “Females are typically more voluble than males when speaking with a same-sex conversational partner of their own choosing [39-41]… Pre-adult females also tend to talk more often than their male counterparts ([42-45]).” Later in the article Locke and Hauser note that: “However, the existence of sex effects in humans is not uniformly accepted. Some investigators consider the claim of greater female volubility a false belief of Western culture, or they attribute it to differing cultural values and expectations for the two sexes [46].”

The research reported in the present work provides a poignant adjustment to the standard viewpoint on sex effects in language since it indicates that speech-like vocalizations (protophones) of the first year are produced much more frequently by boys than by girls.

Additional comments on the results regarding protophone rates reported in the main text

In fact, our report is the first to indicate *any* reliable sex effect in infant vocalization across the first year. A prior much smaller effort showed a trend [47] (see below **Why the greater tendency of boys than girls in producing protophones has not been previously observed**), interestingly also favoring boys, but not statistically reliably. We have never before attempted to test sex effects in infant vocal development because in spite of research across decades in our laboratories, we have never believed we had acquired a sample size large enough to justify the comparison. The reason is that the prior research has been longitudinal and has in the past efforts been conducted microanalytically with very small sample sizes. Based on the much larger sample size of the present work (afforded by the LENA recording method and the random sampling of segments for human coding from those recordings), the results displayed in the Figure in the main text reveal a strong main effect of Sex on protophone rates. At the same time the results show complexity, reflected in the significant interaction of Age by Sex. The boys and girls overlapped in every comparison, so in spite of highly significant statistical tests and large effect sizes, it was not true that boys and girls presented disjunct distributions. These complexities of the outcomes suggest caution in interpretation. Research to test for generalization of the sex differences at larger sample sizes and in additional cultural settings will clearly be important.

Equally important is the possibility that a sex difference favoring boys in vocal rates may apply in the first year but not beyond. We are working currently to assess that possibility with recordings from the current sample and from additional samples that have been acquired by other collaborators.

Cultural and biological factors in the sex difference for protophone volubility

Possible cultural factors in the sex difference reported here. Empirical research of the future should help to isolate factors that influence the sex difference in volubility, if it can be confirmed in future evaluations. One obvious possibility is that cultural factors play a role. Parents might treat boy and girl babies differently in ways that engender different vocalization patterns, for example by talking with them to different extents [34, 48]. Cross-cultural research will clearly be necessary to help tie down any such influences and to illustrate generalizability across circumstances. Our results above on IDS rates suggest weakly that parents in our sample may have talked more to the boys than the girls, although this pattern applied only to the HR infants and did not show the age pattern of sex difference in the HR group that was seen in protophone volubility. That is, the sex difference for protophone volubility dropped consistently across age in the HR group (main text Figure segment C), but the sex difference in IDS showed a much more complex pattern across age, with substantially higher rates for boys both at 0-4.5 mo and at 10.5-13 mo along with lesser differences at intermediate ages.

There are a variety of additional possible cultural variations that could play roles in any sex differences in protophone rates. For example, parents could tend to interrupt boys and girls at different rates, or could tend to listen rather than talk to infant boys and girls to different extents. Such differences could also vary across cultures. The current study does not provide sufficient data to assess such possibilities reliably.

Possible sex differences in activity level as a factor in explaining the sex difference in volubility: Among more biologically-specific possible explanations for the volubility differences across boys and girls, we have suggested physical activity levels, since boys are reported to show higher activity levels than girls [29]. Our data on cry rates in the boys and girls provide weak evidence that might be thought to support a role for activity level since boys in the LR group *did* show slightly higher cry rates overall and substantially higher rates at 0-2.5 months in the LR group. But the numbers of cry utterances were only about 1/6 as high as the numbers of protophones, and 70% of the randomly-selected 5-min segments with infants awake showed no cries at all, while protophones were absent in only 5% of the segments. The relatively low numbers of cries inspire us to doubt the reliability of the trends (non-significant statistically in any case) and leave the question of possible differences in cry rates across boys and girls relatively open. Obviously, it will be useful to collect a larger sample, including more cries across many more infants, as well as data on other aspects of activity levels (finger movements, limb movements, locomotion, etc.) across sexes and across cultures.

Reasons to consider naturally selected "fitness signaling" as a basis for high protophone rates in human infants: Another biologically-oriented speculation for the sex difference in protophone rates requires background explanation involving decades of research on the protophones that has until now not focused on possible sex differences. This section provides that background, followed by a section with speculation about why we think the fitness-signaling hypothesis, if it proves viable, may provide a basis for explaining the sex difference.

The fitness-signaling hypothesis is founded in a proposed explanation for the very high rate of protophone production in the infants that have both studied both in the present work and in a variety of prior research projects; the rate is now estimated to be about 3500 protophones per day [13] for typically developing infants. It is important to emphasize that *very high rates in this prior research have applied to both boys and girls*, although until now we have not had a sufficiently large sample to reliably estimate possible sex differences. Given the present outcome, we can estimate that boys produce more than five protophones per min, while girls produce more than four per min. Both these rates are vastly higher than the rates that have been empirically estimated for our closest phylogenetic relatives, who appear to produce some acoustically protophone-like sounds in infancy, but apparently less than one-tenth as many as either human boys or human girls (see data and review of related studies in Oller et al. [8]).

The background studies suggest substantial generality to the high protophone rates in human infants:

1) with SES ranging from high to very low in the USA [49, 50], where only about a 20% higher rate has been found in infants with higher SES;
2) born prematurely or at term [13, 50], where no significant differences have been observed;
3) in several cross-cultural investigations, none of which has turned up notable volubility differences between the languages/cultures examined in
    (a) Spanish- and English-speaking homes in Miami [51];
    (b) English- and Mandarin/Southern Min-speaking homes in Memphis and in Taiwan [52]; or
    (c) English- and Lebanese Arabic-speaking families in Memphis and in Lebanon [53]; and
4) with a variety of disorders of communication where volubility differences between typically developing infants and infants with disorders have been (as with the sex differences) very small by comparison with the massive differences between human infants and our closest phylogenetic relatives, bonobos and chimpanzees. These studies have included infants with
    (a) Down Syndrome [54];
    (b) autism [21];
    (c) fragile X syndrome [55]; and
    (d) profound deafness [56-59].

The case of profound deafness at birth is perhaps the most instructive on the above list. A long-standing expectation, widely cited, is that deaf infants would vocalize little [60], but results of empirical research simply do not conform to that expectation. Even infants with essentially no auditory input produce protophones at rates similar to those of hearing human infants, a pattern we interpret as consistent with the notion that the production of human infant protophones is driven heavily by internal, endogenous factors. Nonetheless, we have withheld any suggestion that very high protophone rates in human infants are universal, pending studies on protophone rates in very low technology cultures. We know of no study yet published comparing infant protophone volubility across cultures where one is relatively affluent (first world) and the other is very poor by common standard-of-living indices (third world). Research on IDS in third-world cultures suggests it is (at least in some cases) considerably lower than in cultures benefiting from high technology (see data and review of prior research in Cristia et al. [61]). Surprisingly perhaps, the same studies have not addressed infant volubility quantitatively.

While there are vast differences between the human infant and other ape infants (bonobos and chimpanzees having been best studied, but still only to a very limited extent) in relative production of protophones or protophone-like sounds, there are other sound types that show greater similarity across the apes. We share *cry/scream* utterance types that are fairly easily identified as such across humans and other apes from a functional standpoint, and in all the primate species, these occur commonly when infants are judged to be in distress (see review in [8] and [62]). On the other hand, the human infants we have observed have produced thousands of *protophones* daily even from the first month, the vast majority of which have been produced with no sign of distress. In fact, only about 15% of protophones in our research have been deemed functionally “negative” [6, 7]; we sometimes term these protophones “complaints”, although they are not coded as “cries”. The vast majority of the remaining protophones have not only been produced in apparent comfort (showing no sign of facial or auditory negativity), but most of them have been *directed toward no person*. Rather, they seemed to have been produced in non-interactive vocal play and vocal exploration [63]. In general the protophones do not appear to be elicited by social interaction, but instead appear to be motivated from within the infant. Even human infants alone in a room produced protophones in apparent comfort at a rate of ~4/min [13].

These are indications of endogenous vocal motivation for the protophones produced by the human infant. The 3 bonobo infants that have been studied in direct comparison (in the only such study that exists [8]), did not clearly produce *any* protophones—they did clearly produce screams (clear distress sounds), laughter, and some low intensity complaint sounds, which were interpreted as “protophone-like”. The most protophone-like sounds (from an acoustic standpoint) the bonobo infants produced were *never* judged to be produced playfully or exploratorily, while virtually all that *could* be judged for affect were judged to be negative (complaints, pleas for help, etc.), leaving open the interpretation that they were low intensity versions of screams rather than being protophones of any sort. All in all, it seems clear that the human infant is inclined to explore the vocal capacity actively for no interactive purpose (although interaction is of course possible) and usually in the absence of any negative emotion, while there is no evidence that any other ape infant (or adult) produces vocalization in a similar way or to a similar extent. More research is needed of course, but at the very least it appears the human infant is far more inclined to vocalize exploratorily than other ape infants.

These conclusions suggest to us that there has been selection pressure on hominin infants to vocalize exploratorily, and combined with the above reasoning, we deem it sensible to suggest that hominin infants have been naturally selected to produce protophones. But what force of natural selection could account for this difference between hominins and other apes?

There is only one proposal on the table, as far as we know. The existing proposal suggests there has been natural selection pressure in the human lineage on fitness signaling due to altriciality of the human/hominin infant, who faces a long infancy and childhood during which high parental investment is critical to survival [64, 65]. Other apes can be said to be less altricial, that is to have relatively shorter periods of helplessness [66]. Human infants under this proposal are thought to have been selected to be instinctively motivated to explore the vocal capability, and caregivers are thought to have been selected to *notice* these vocalizations, even when they are not directed toward them, as indicators of infant wellness and worthiness of long term investment.

It is important to recognize that the protophones are *usually produced in comfort*, and consequently are not treated by caregivers as expressions of need (as for example cries are). Instead, when they are heard by caregivers, they supply information about the wellness of the infant, precisely the kind of information that might be thought useful in making caregiving investment decisions. The ancient hominin infant who produced copious comfort signaling protophones thus may have (in accord with the reasoning) increased the chances of being nurtured and also of not being abandoned.

It appears that the importance of the protophones as indicators of normal development is high for all infants, but perhaps particularly high for infants at risk for significant language-related disorders, and perhaps it has been so in deep time. Our research has been directed in part to comparing protophone production in infants at high risk for autism with a control group of infants at low risk. As will be indicated below, boys can also be considered generally at higher risk for death than girls, especially in the first year.

A practice hypothesis about the origin of the protophones? It is sensible to propose that the protophones constitute practice with the vocal mechanism, yielding benefits for the later acquisition of language. And indeed in modern times, it does appear the protophones constitute a kind of practice that is relevant to the development of the capacity for speech in much the same way that practice in crawling or scooting is relevant to the development of later locomotory skills such a walking and running. But there is reason to resist hypothesizing that the origin of the protophones has its foundations in selection pressure on practice for a speech capacity, given that speech would not have existed at the time the first exploratory vocalizations would have occurred in hominins. Natural selection must target capabilities that have survival/reproduction advantages at the point at which they begin to emerge. Evolutionary-developmental biology (evo-devo) [67-72] posits a fundamental principle that natural selection preferentially targets developmental processes—no capacity can be evolved if it cannot be developed. The steps of evolution have to build, in accord with this reasoning, on existing capabilities that must be molded in development. To posit practice for speech as the ultimate source of selection pressure on protophone production thus would seem to have the origin story backwards. Flexible vocal capacity must, we presume, have been targeted first, and once established, could have supplied a foundation from which more speech-like activities could have been developed and evolved.

Why fitness signaling has not been a focus of research in mammalian infant development: Research on fitness signaling has focused heavily on songbirds, where territoriality and mating clearly often involve vocal displays of fitness [73-76]. The focus of such research is on mature birds. An emerging literature on *infant* fitness signaling [77-80] has often been preoccupied with the “handicap principle” [81, 82] and with issues regarding equilibria between conflicting parent and infant interests, to the exclusion of consideration of fitness signaling that could profit both sender and receiver.

However, there has been some attention paid to the idea that “playful” behavior in the young of primates, including humans constitutes a fitness signal [83, 84]. We and Locke [64, 66, 85, 86], have focused on the protophones as presumably very low-cost signals, often executed playfully, that may *not* involve significant conflicting parent and infant interests (and see Fitch [87], who presents a related argument regarding the origin of language). Instead, protophones appear to provide information that is beneficial on both sides—the infant seeks higher investment through comfortable protophone production (not necessarily intentionally), and the parent uses the signals to make judicious investment decisions about each infant. The vast numbers of protophones produced by human infants (both boys and girls) call for an explanation, especially since the evidence suggests no other ape produces such signals [8, 88, 89]. We are not wedded to a fitness signaling explanation, but it seems a plausible explanation for the vast number of protophones in human infants (we are open to other suggestions about the roots of the tendency to vocalize so frequently in the absence of obvious immediate benefits).

It is a matter of some surprise to us that the issue of possible fitness signaling in the young of mammals has not been more thoroughly considered in behavioral biology. All mammals are dependent on parental care as infants, and consequently it seems reasonable to propose that all are under selection pressure to provide signals to advertise their fitness and secure parental investment. Competition among the young of mammals for parental investment would seem to be a foregone conclusion, and consequently competition for that investment through advertisement seems likely. Mammal parents in general, in accord with this reasoning, would also be selected to recognize, evaluate, and use such infant signals as a means of making effective investment decisions.

Importantly, this reasoning does not preclude conflicts between infant need signaling and parental investment decisions. Clearly there exist inherent conflicts between parents and infants, where cost equilibria are developed between the needs of each. But these conflicts occur *between* any individual parent and any individual infant and pertain to signals of need, such as crying. In contrast, infant fitness signals such as protophones do not usually constitute expressions of immediate need; to the extent that protophones involve competition, it is not *between* parents and infants, but *among infants*, who can be viewed as being in competition with each other for parental investment. The protophones supply information useful to parents in making these investment decisions with regard to various infants. We reason that parental investment is guided by both infant signals of need such as crying (and to that extent, equilibrium must be developed between parental and infant needs) and by information from fitness indicators (including protophones and other fitness indicating characteristics of each infant) helping the parent decide (whether consciously or unconsciously) how much to invest in each infant.

A sex difference hypothesis founded in the fitness signaling hypothesis: The notion that fitness signaling may drive greater production of protophones in boys than in girls starts from the reasoning above, proposing a fitness signaling function for the protophones. The reasoning is at least loosely consistent with the well-known biological principle of fitness signaling. Thousands of species (including at least songbirds, bats, frogs, cetaceans, and pinnipeds) use vocalization in fitness displays for mating and territorial defense, with much more extensive male than female activity [90].

The proposal that hominin infants have long used protophone-like vocalizations as signals of well-being represents an extension of the well-documented and genetically-determined tendency for animals to capitalize on vocalization as an opportunity to advertise to further their interests. It also invokes the notion that parents make informed decisions regarding their investments in offspring [91]. The apparent exaggeration of the sex effect in the earliest months for the HR infants further hints at genetic foundations for fitness signaling through vocalization, which may be increased in frequency among infants at high risk for death or disability. Importantly for our reasoning, boys in general are reported to be more at risk for death in the first year than girls [1-3].

Consequently among the various possible both cultural and biological explanations for the apparent greater production of protophones by infant boys than girls, we see the idea of fitness signaling as providing a viable option. Boys can be thought to be under higher selection pressure for fitness signaling than girls, and especially in the first year, because boys are presumably at high risk especially in the first year. We are now seeking to make comparisons based on all-day recordings between boys and girls on vocalization rates *beyond* the first year.

Of course the optimal explanation for sex differences in protophone rates may lie in a combination of the several cultural and biological factors that have been considered here, as well as others that have not yet been specified.

On testability of hypotheses about both cultural and biological influences on human infant signaling and possible sex differences: In the long run, we need additional both observational and experimental studies to help evaluate the viability of the fitness signaling idea and other proposals (both cultural and biological) to explain protophone rates in human infants and to explain the sex difference, if it holds up in future research. For example, a broadscale study of parent-infant interaction could assess a wide variety of possible investment types (nourishment provision, protection from harm, willingness to assist in travel, communicative attentiveness, etc.) and quantities in association with amount of protophone and/or cry production. Such research could surely provide relevant evidence of the possible effectiveness of either kind of sound as a fitness signal. Furthermore, experimental studies could assess physiological reactions of parents or potential parents in circumstances where they listen to infant vocalizations, both protophones and cries. There are surely many additional ways to evaluate fitness signaling empirically, as well as to evaluate other proposals regarding possible reasons for high protophone rates and for possible sex differences.

Why the greater tendency of boys than girls in producing protophones has not been previously observed

Based on the present results, it appears the tendency of boys to produce more protophones than girls is very robust by comparison with other reported effects of sex, where overwhelmingly, girls have been reported to show better language and language-related skills. The effect size found in the present work favors the boys with a Cohen's d of 0.89, while reported effects favoring in girls in prior studies have been generally <0.2. Given the robustness of the effect, it is reasonable to ask why it has not been previously noticed.

First, empirical research on infant vocal development, focused on infant vocalizations themselves, protophones as well as cries, and not instead focusing on parent-infant interaction, has been infrequent. Until the last few years, there has been very little research estimating the rate of infant vocalization. Much of the relevant research, supplying counts of vocal events in human infants, especially in longitudinal work, has come from our own laboratories or those of our close collaborators [16].

There is, however, one prior study on sex differences in infant vocal rates that deserves comment. Sung et al. evaluated recordings of 30 infants (15 boys and 15 girls) with their parents in their homes [47]. They counted protophones, and although boys produced more protophones than girls overall, the differences were not statistically significant and varied across the three ages into which the authors categorized the observations. The major difference between Sung et al. and the present study is sheer size. They evaluated approximately 2700 minutes of recorded data. The present study evaluated over 90,000 minutes, 33 times more data, and the present study involved more than three times as many infants. There are other differences that may be of note, particularly that our recordings were made in the absence of any academic observers. Further, our recordings did not require the parents to interact with the infants (indeed, the vast majority of the time, infants were not engaged in interaction in our recordings), whereas Sung et al. selected 18 five-minute periods for each infant during which the mother was specifically engaged in interaction with the infant. Our investigation thus worked with data from the natural environment of the infant at home, whereas the Sung et al. data were less naturalistic.

Author contributions

D.K.O., G.R. and U.G. drafted the main text and Supplemental Information. D.K.O prepared Figures and Table in both the main text and the Supplemental Information. D.K.O., G.R., and E.H.B designed the research. D.D.B. supervised statistical analyses and conducted the GEE analyses. G.R. supervised and coordinated recordings. D.K.O., E.R.B., Y.J., H.Y., and H.L. trained coders and supervised the coding process. All authors reviewed the manuscript and helped refine the writing

Supplemental References

S1. Mage, D.T., and Donner, M. (2004). The X-linkage hypotheses for SIDS and the male excess in infant mortality. Medical Hypotheses *62*, 564-567.

S2. Töro, K., Sawaguchi, T., Sawaguchi, A., Rózsa, S., and Sótonyi, P. (2001). Comparative analysis of differences by gender in sudden infant death syndrome in Hungary and Japan. Forensic Sci Int. *118*, 9-15.

S3. Zhao, D., Zou, L., Lei, X., and Zhang, Y. (2017). Gender Differences in Infant Mortality and Neonatal Morbidity in Mixed-Gender Twins. Scientific Reports *7*, 8736.

S4. Oller, D.K., Niyogi, P., Gray, S., Richards, J.A., Gilkerson, J., Xu, D., Yapanel, U., and Warren, S.F. (2010). Automated Vocal Analysis of Naturalistic Recordings from Children with Autism, Language Delay and Typical Development. Proceedings of the National Academy of Sciences *107*, 13354-13359.

S5. Oller, D.K. (2000). The Emergence of the Speech Capacity, (Mahwah, NJ: Lawrence Erlbaum Associates).

S6. Jhang, Y., and Oller, D.K. (2017). Emergence of Functional Flexibility in Infant Vocalizations of the First 3 Months. Frontiers in Psychology *8*, 300.

S7. Oller, D.K., Buder, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L., and Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. Proceedings of the National Academy of Sciences *110*, 6318-6632.

S8. Oller, D.K., Griebel, U., Iyer, S.N., Jhang, Y., Warlaumont, A.S., Dale, R., and Call, J. (2019). Language origin seen in spontaneous and interactive vocal rate of human and bonobo infants. Frontiers Psychology *10*.

S9. Kojima, S. (2003). A Search for the Origins of Human Speech, (Kyodai Kaikan: Kyoto University Press).

S10. Laporte, M.N.C., and Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. Animal Behaviour *80*, 467-473.

S11. Yoo, H., Buder, E.H., Bowman, D.D., Bidelman, G.M., and Oller, D.K. (2019). Acoustic Correlates and Adult Perceptions of Distress in Infant Speech-Like Vocalizations and Cries. Frontiers in Psychology *10*, 1154.

S12. Jhang, Y., Franklin, B., Ramsdell, H.L., and 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*.

S13. Oller, D.K., Caskey, M., Yoo, H., Bene, E.R., Jhang, Y., Lee, C.-C., Bowman, D.D., Long, H.L., Buder, E.H., and Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. Scientific Reports. *9*, 14734.

S14. Lynch, M.P., Oller, D.K., Steffens, M.L., and Buder, E.H. (1995). Phrasing in prelinguistic vocalizations. Developmental Psychobiology *28*, 3-23.

S15. Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In Child phonology, Vol 1: Production, G. Yeni-Komshian, J. Kavanagh and C. Ferguson, eds. (New York: Academic Press), pp. 93-112.

S16. Iyer, S.N., Ertmer, D.J., and Stark, R.E. (2006). Assessing vocal development in infants and toddlers. Clinical Linguistics & Phonetics *20*, 351–369.

S17. Koopmans-van Beinum, F.J., Clement, C.J., and van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? Developmental Science *4*, 61-70.

S18. 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 *4*, 158-164.

S19. Papoušek, M. (1994). Vom ersten Schrei zum ersten Wort: Anfänge der Sprachentwickelung in der vorsprachlichen Kommunikation, (Bern: Verlag Hans Huber).

S20. Molemans, I., Van den Berg, R., Van Severen, L., and Gillis, S. (2011). How to measure the onset of babbling reliably. Journal of Child Language, 1-30.

S21. Patten, E., Belardi, K., Baranek, G.T., Watson, L.R., Labban, J.D., and Oller, D.K. (2014). Vocal patterns in infants with Autism Spectrum Disorder: Canonical babbling status and vocalization frequency. Journal of Autism and Developmental Disabilities *44*, 2413-2428.

S22. Delgado, R.E., Buder, E.H., and Oller, D.K. (2010). AACT (Action Analysis Coding and Training). (Miami, FL: Intelligent Hearing Systems).

S23. Liang, K.-Y., and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. Biometrika *73*, 13-22.

S24. Borjon, J.I., and Ghazanfar, A.A. (2014). Convergent Evolution of Vocal Cooperation without Convergent Evolution of Brain Size. Brain Behav Evol 84, 93–102.

S25. Elowson, A.M., Snowdon, C.T., and Lazaro-Perea, C. (1998). 'Babbling' and social context in infant monkeys: parallels to human infants. Trends in cognitive sciences *2*, 31-37.

S26. Snowdon, C.T., and Elowson, A.M. (1999). Pygmy marmosets modify call structure when paired. Ethology *105*, 893-908.

S27. Takahashi, D.Y., Fenley, A.R., Teramoto, Y., Narayanan, D.Z., Borjon, J.I., Holmes, P., and Ghazanfar, A.A. (2015). The developmental dynamics of marmoset monkey vocal production. Science *349*, 734-738.

S28. Takahashi, D.Y., Narayanan, D.Z., and Ghazanfar, A.A. (2013). Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. Current Biology 23, 2162–2168.

S29. Campbell, D.W., and Eaton, W.O. (1999). Sex differences in the activity level of infants. Infant and Child Development *8*, 1-17.

S30. Gilkerson, J., Richards, J.A., Warren, S.F., Oller, D.K., Russo, R., and Vohr, B.R. (2018). Language Experience in the Second Year of Life and Language Outcomes in Late Childhood. Pediatrics *142*, 1-11.

S31. Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L.V. (2010). Sources of variability in children's language growth. Cognitive psychology *61*, 343-365.

S32. Kitamura, C., Thanavishuth, Burnhama, D., and Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. Infant Behavior & Development *24*, 372–392.

S33. Schwab, J.F., and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. Wiley Interdiscip Rev Cogn Sci *7*, 264-275.

S34. Bakeman, R., and Brown, J.V. (1977). Behavior dialogues: An approach to the assessment of mother-infant interaction. Child Development *48*, 195-203.

S35. Bloom, K., Moore-Schoenmakers, K., and Masataka, N. (1999). Nasality of infant vocalizations determines gender bias in adult favorability ratings. Journal of Nonverbal Behavior *23*, 219-236.

S36. Fausto-Sterling, A., Crews, D., Sung, J., García-Coll, C., and Seifer, R. (2015). Multimodal sex-related differences in infant and in infant-directed maternal behaviors during months three through twelve of development. Developmental psychology *51*, 1351-1366.

S37. Johnson, K., Caskey, M., Rand, K., Tucker, R., and Vohr, B.R. (2014). Gender Differences in Adult-Infant Communication in the First Months of Life. Pediatrics *134*, e1603–e1610.

S38. Locke, J.L., and Hauser, M. (1999). Sex and Status Effects on Primate Volubility: Clues to the Origin of Vocal Languages? Evolution and Human Behavior *20*, 151–158.

S39. Dabbs, J.M., and Ruback, R.B. (1984). Vocal patterns in male and female groups. Personality and Social Psychology Bulletin *10*, 518-525.

S40. Ickes, W., and Barnes, R.D. (1977). The Role of Sex and Self-Monitoring in Unstructured Dyadic Interactions. Journal of Personality and Social Psychology *35*, 315-330.

S41. Street, R.L., and Murphy, T.J. (1987). Interpersonal orientation and speech behavior. Communication Monographs *54*, 42–62.

S42. Jormakka, L. (1976). The behaviour of children during a first encounter. Scandanavian Journal of Psychology *17*, 15-22.

S43. Larson, R.W., Richards, M.H., Moneta, G., Holmbeck, G., and Duckett, E. (1996). Changes in adolescents' daily interactions with their families from ages 10 to 18: disengagement and transformation. Developmental Psychology *32*, 744–754.

S44. Raffaelli, M., and Duckett, E. (1989). "We were just talking…": Conversations in early adolescence. Journal of Youth and Adolescence *18*, 567-582.

S45. Smith, P.K., and Connolly, K. (1972). Patterns of play and social interaction in pre-school children. In Ethological studies of child behaviour. (Oxford, England: Cambridge U. Press), pp. x, 400-x, 400.

S46. James, D., and Drakich, J. (1993). Understanding gender differences in amount of talk: A critical review of research. In Gender and conversational interaction. (New York, NY, US: Oxford University Press), pp. 281-312.

S47. Sung, J., Fausto-Sterling, A., Garcia Coll, C., and Seifer, R. (2013). The Dynamics of Age and Sex in the Development of Mother–Infant Vocal Communication Between 3 and 11 Months. Infancy, 1–24.

S48. Johnson, K., Caskey, M., Rand, K., Tucker, R., and Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. Pediatrics *134*, e1603-1610.

S49. Oller, D.K., Eilers, R.E., Basinger, D., Steffens, M.L., and Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. First Lang *15*, 167-188.

S50. Eilers, R.E., Oller, D.K., Levine, S., Basinger, D., Lynch, M.P., and Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. Infant Behavior and Development *16*, 297-315.

S51. Oller, D.K., Eilers, R.E., Urbano, R., and Cobo-Lewis, A.B. (1997). Development of precursors to speech in infants exposed to two languages. Journal of Child Language *27*, 407-425.

S52. Lee, C.-C., Jhang, Y., Relyea, G., and Oller, D.K. (2017). Subtlety of ambient-language effects in babbling: A study of English- and Chinese-learning infants at 8, 10, and 12 months. Language Learning and Development *13*, 100-126.

S53. Farran, L.K., Yoo, H., Lee, C.-C., Bowman, D.D., and Oller, D.K. (2019). Temporal Coordination in Mother–Infant Vocal Interaction: A Cross-Cultural Comparison. Frontiers in Psychology *10*.

S54. Lynch, M.P., Oller, D.K., Steffens, M.L., Levine, S.L., Basinger, D., and Umbel, V. (1995). The onset of speech-like vocalizations in infants with Down syndrome. American Journal of Mental Retardation *100*, 68-86.

S55. Belardi, K.M., Watson, L.R., Faldowski, R., Baranek, G.T., Crais, B., Patten, E., Oller, D.K., Hazlett, H., and McComish, C. (2017). A Retrospective Video Analysis of Canonical Babbling and Volubility in Infants with Fragile X Syndrome at 9 -12 Months of Age. Journal of Autism and Developmental Disabilities *47*, 1193– 1206.

S56. Van den Dikkenberg-Pot, I., Koopmans-van Beinum, F., and Clement, C. (1998). Influence of lack of auditory speech perception of sound productions of deaf infants. Proceedings of the Institute of Phonetic Sciences, University of Amsterdam *22*, 47-60.

S57. Iyer, S.N., and Oller, D.K. (2008). Prelinguistic vocal development in infants with typical hearing and infants with severe-to-profound hearing loss. Volta Review *108*, 115-138.

S58. Oller, D.K., and Eilers, R.E. (1988). The role of audition in infant babbling. Child Development *59*, 441-449.

S59. Eilers, R.E., and Oller, D.K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. Journal of Pediatrics *124*, 199-203.

S60. Lenneberg, E. (1967). Biological foundations of language, (New York: John Wiley & Sons, Inc.).

S61. Cristia, A., Dupoux, E., Gurven, M., and Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. Child Development, *90*, 759-773.

S62. Bard, K.A. (2000). Crying in infant primates: Insights into the development of crying in chimpanzees. In Crying as a sign, a sympton, and a signal: Developmental and clinical aspects of early crying behavior, R. Barr, B. Hopkins and J. Green, eds. (London, UK: MacKeith Press), pp. 157 - 175.

S63. Long, H.L., Bowman, D.D., Yoo, H., Burkhardt-Reed, M.M., Bene, E.R., and Oller, D.K. (in submission). Social and non-social functions of infant vocalizations.

S64. Locke, J.L. (2006). Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. Human Nature *17*, 155-168.

S65. Oller, D.K., and Griebel, U. (2005). Contextual freedom in human infant vocalization and the evolution of language. In Evolutionary Perspectives on Human Development, R. Burgess and K. MacDonald, eds. (Thousand Oaks, CA: Sage Publications), pp. 135-166.

S66. Locke, J.L., and Bogin, B. (2006). Language and life history: A new perspective on the evolution and development of linguistic communication. Behavioral & Brain Sciences *29*, 259-325.

S67. 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. Phil. Trans. R. Soc. B *366*, 2055-2180.

S68. Carroll, S.B. (2005). Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom, (New York: W. W. Norton).

S69. Gottlieb, G. (2002). Developmental–behavioral initiation of evolutionary change. Psychological Review *109*, 211–218.

S70. Locke, J.L. (2009). Evolutionary developmental linguistics: Naturalization of the faculty of language. Language Sciences *31*, 33–59.

S71. Müller, G.B., and Newman, S.A. (2003). Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology, (Cambridge MA: MIT Press).

S72. Newman, S.A., and Müller, G.B. (2000). Epigenetic mechanisms of character origination. J Exp Zool B Mol Dev Evol *288*, 304-317.

S73. Catchpole, C.K. (1982). The evolution of bird sounds in relation to mating and spacing behavior. In Acoustic Communication in Birds: Vol 1: Production, Perception and Design Features of Sounds, D.E. Kroodsma and E.H. Miller, eds. (New York, NY: Academic Press), pp. 297-319.

S74. Catchpole, C.K., and Slater, P.J.B. (2003). Bird song: Biological themes and variations, (Cambridge: Cambridge University Press).

S75. Kroodsma, D.E. (1999). Making ecological sense of song development. In The design of animal communication, M.D. Hauser and M. Konishi, eds. (Cambridge, MA: MIT Press), pp. 319-342.

S76. Marler, P., and Slabbekorn, H. (2004). Nature's Music: The Science of Birdsong, (Amsterdam: Elsevier Academic Press).

S77. Godfray, H.C.J. (1991). Signalling of need by offspring to their parents. Nature *352*, 328-330.

S78. Lachmann, M., Számadó, S., and Bergstrom, C.T. (2001). Cost and conflict in animal signals and human language. Proceedings of the National Academy of Sciences *98*, 13189-13194.

S79. Penn, D.J., and Számadó, S. (2020). The Handicap Principle: how an erroneous hypothesis became a scientific principle. Biological Reviews *95*, 267-290.

S80. Rodríguez-Gironés, M.A., Cotton, P.A., and Kacelnik, A. (1996). The evolution of begging: Signaling and sibling competition. Proceedings of the National Academy of Sciences *93*, 14637-14641.

S81. Grafen, A. (1990). Biological Signals as Handicaps. Journal of theoretical biology *144*, 517-546

S82. Zahavi, A. (1975). Mate selection--a selection for a handicap. Journal of theoretical biology *53*, 205-214.

S83. Lafreniere, P.J. (2011). Evolutionary Functions of Social Play: Life Histories, Sex Differences, and Emotion Regulation. American Journal of Play *3*, 464-488.

S84. Mason, W.A. (1965). The social development of monkeys and apes. In Primate Behavior: Field studies of monkeys and apes, I. DeVore, ed. (New York: Holt, Rinehart & Winston), pp. 514–544.

S85. Oller, D.K., and Griebel, U. (2006). How the language capacity was naturally selected: Altriciality and long immaturity. Commentary on Locke, J. and Bogin, B: Language and life history: A new perspective on the evolution and development of linguistic communication. Behavioral and Brain Sciences. *29*, 293-294.

S86. Oller, D.K., Griebel, U., and Warlaumont, A.S. (2016). Vocal development as a guide to modeling the evolution of language Topics in Cognitive Science (topiCS), Special Issue: New Frontiers in Language Evolution and Development, Editor, Wayne D. Gray, Special Issue Editors, D. Kimbrough Oller, Rick Dale, and Ulrike Griebel *8*, 382-392.

S87. Fitch, W.T. (2004). Evolving honest communication systems: kin selection and "mother tongues". In The Evolution of Communication Systems: A Comparative Approach, D.K. Oller and U. Griebel, eds. (Cambridge, MA: MIT Press), pp. 275-296.

S88. Hauser, M. (1996). The evolution of communication, (Cambridge, MA: MIT).

S89. Owren, M.J., Amoss, R.T., and Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. American Journal of Primatology *73*, 530-544.

S90. Griebel, U., Pepperberg, I.M., and Oller, D.K. (2016). Developmental plasticity and language: A comparative perspective. Topics in Cognitive Science (topiCS) *8*, 435-445.

S91. Trivers, R.L. (1972). Parental investment and sexual selection. In Sexual selection and the descent of man, 1871-1971, B. Campbell, ed. (Chicago, IL: Aldine), pp. 136-179.