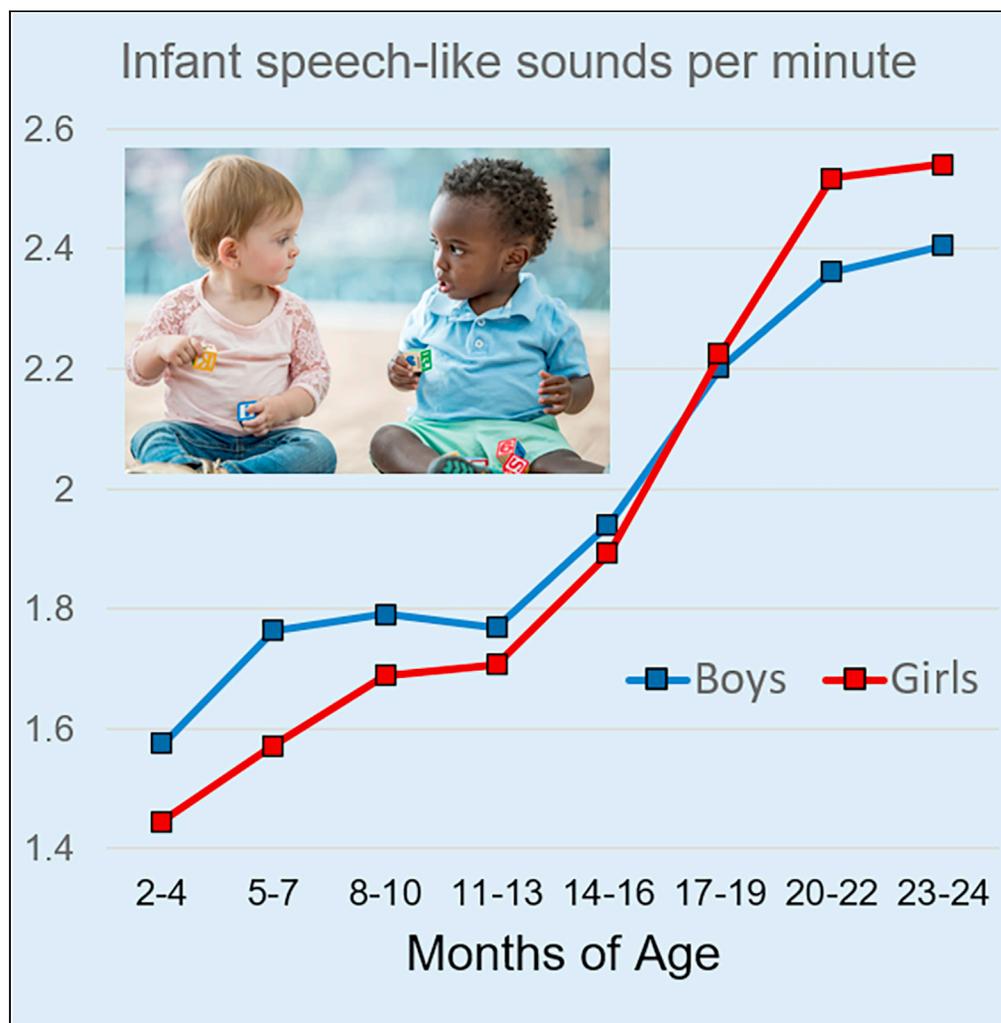Article

Sex differences in infant vocalization and the origin of language

D. Kimbrough Oller, Jill Gilkerson, Jeffrey A. Richards, ..., Jane A. Brown, Hyunjoo Yoo, Steven F. Warren

koller@memphis.edu

Highlights

Females are widely believed to have an advantage over males in language

Yet first year boys produced more speech-like vocalizations than girls in recent work

In 450,000 h of recording, we also found boys more voluble than girls

But higher rate of male vocalization occurred in the first but not the second year of life

iScience

Article

Sex differences in infant vocalization and the origin of language

D. Kimbrough Oller,[1,2,3,7,*] Jill Gilkerson,[4] Jeffrey A. Richards,[4] Steve Hannon,[4] Ulrike Griebel,[1,2,3] Dale D. Bowman,[1] Jane A. Brown,[1,2] Hyunjoo Yoo,[5] and Steven F. Warren[6]

SUMMARY

Seeking to discern the earliest sex differences in language-related activities, our focus is vocal activity in the first two years of life, following up on recent research that unexpectedly showed boys produced significantly more speech-like vocalizations (protophones) than girls during the first year of life. We now bring a much larger body of data to bear on the comparison of early sex differences in vocalization, data based on automated analysis of all-day recordings of infants in their homes. The new evidence, like that of the prior study, also suggests boys produce more protophones than girls in the first year and offers additional basis for informed speculation about biological reasons for these differences. More broadly, the work offers a basis for informed speculations about foundations of language that we propose to have evolved in our distant hominin ancestors, foundations also required in early vocal development of modern human infants.

INTRODUCTION

In a recent article, hereafter CB-2020,[1] it was reported that male infants produced a significantly higher rate of speech-like vocalizations (termed "protophones") than female infants across the first year of life. The finding was surprising given that females are generally reported to have a discernible advantage over males in language.[2] Irrespective of sex, the ability and tendency of human infants to produce a wide range of protophones at very high rates across the first year has been argued to be a necessary foundation for the development of vocal language,[3–5] so a sex-based difference in rate of production cries out for explanation. CB-2020 argued tentatively for a fundamental biological explanation, invoking the fact that baby boys are more vulnerable to death in the first year than baby girls,[6] and consequently that boys may be now, and may have been in the distant hominin past, under greater natural selection pressure than girls to signal their wellness to caregivers through comfortable vocalization. According to the reasoning, the altricial (born helpless) human infant, facing years of dependency, needs long-term commitment from caregivers; protophones produced in comfort are posited to constitute fitness signals; hence, we refer to this framework of thought as the Fitness Signaling Theory.[7,8] Caregivers are thus reasoned to respond to hearing protophones produced in comfort as evidence of wellness, and whether tacitly or consciously, they are assumed to implement selection pressure by investing differentially in infants based on the wellness evidence manifest in the protophones. Continuing the reasoning, since boys are more vulnerable to dying in the first year,[6] they were reasoned in CB-2020 to be under greater selection pressure to signal their fitness than girls, and thus, to produce more protophones.

CB-2020 was based on human coding of randomly selected segments from longitudinal all-day recordings of 65 boys and 35 girls across the first year. The present work re-addresses the empirical claim that boy babies produce more protophones than girl babies, relying on the availability of an enormous database of automatically categorized protophones from all-day recordings of thousands of infants, from all over the USA, studied semi-longitudinally across the first two years of life. In addition, exploiting this massive dataset provides the opportunity to delve more deeply into how the Fitness Signaling Theory may offer a biological basis to explain the reported early sex difference in protophone volubility.

The common assertion of a female language advantage can be traced back to the middle of the last century.[9–12] Surprisingly, the relevant empirical data yield a complex picture where most studies have actually shown no statistically reliable differences between the sexes, and where many studies, even with large sample sizes, have reported no female advantage at all.[2] In fact, in a number of prominent studies, the results

[1]University of Memphis, School of Communication Sciences and Disorders, Memphis, TN 38152, USA

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

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

[4]LENA Foundation, Louisville, CO 80027, USA

[5]University of Alabama, Communicative Disorders, Tuscaloosa, AL 35487, USA

[6]University of Kansas, Speech-Language-Hearing: Sciences and Disorders, Lawrence, KS 66045, USA

[7]Lead contact

*Correspondence: koller@memphis.edu

https://doi.org/10.1016/j.isci.2023.106884

have suggested the opposite. In the largest study on record,[13] where the verbal scores on the Stanford Achievement Test for nearly a million graduating high school seniors in 1985 were compared, males came out on top, by a small but statistically reliable margin. Still, the opinion persists that females have a statistically discernible advantage over males in language,[14] an opinion supported by most investigations that have yielded reliable results and by continuing research of high quality.[15–17] Since both sexes must use language extensively, it seems unsurprising that whatever differences do occur must be limited to show small effect sizes.

The hundreds of studies that have been conducted on sex and language appear to differ in outcomes in part because they have evaluated different features of language and/or different circumstances of its usage. Leaper and colleagues[18,19] have helped clarify the diverse findings with meta-analyses distinguishing among three areas of comparison: 1) affiliative language, roughly the tendency to make social connections with language, 2) assertive language, roughly the tendency to display dominance or self-importance with language, and 3) talkativeness, in our own terminology "volubility". In general, the findings suggest females use more and show advantages in production of affiliative language, while males use more and show advantages in assertive language. Talkativeness tends to be higher for males in adulthood but for females in childhood. Yet even this summary glosses over a multitude of complexities involving modulating variables explored by Leaper and colleagues, who concur with authors of the prior most important meta-analysis[2] by emphasizing that the significant differences reported tend to have small to very small effect sizes (Cohen's $d < 0.2$) and tend sometimes to reverse in male or female advantage from one circumstance of observation to another. Furthermore, sex and language research has shown results that can differ diametrically based on the composition of speaker pairs (same-sex or mixed, peer-with-peer or caregiver-with-child, and so on) or group sizes.

Continuing widespread interest in sex and language appears to be motivated by fundamental questions about the biology of human communication as affected by sexual dimorphism and its sensitivity to cultural factors that influence gender identity and gender roles across the lifespan, as well as gender-differentiated usage of language in varying contexts. The increasing modern emphasis on a distinction between sex and gender, with significant numbers of individuals asserting gender not corresponding to their biological sex, especially during or as they approach adolescence, intensifies the interest in the potential roles of both sex and gender in language usage.[20,21] The existing literature offers theoretical perspectives invoking both biological and cultural factors influencing male and female roles across a variety of societies and evaluating how those roles create different patterns of language use.[22,23] That sex differences in language-related vocalization may occur in the very first year of life is intriguing in part because it may provide an opportunity to provide perspective on language foundations.

Research by some of the present authors has long focused on illuminating the origins of language,[4,24–27] and consequently the interest in sex differences has led us to address the earliest manifestations of language-related differences that can be monitored. These interests tend to align theoretically with evo-devo, i.e., evolutionary-developmental biology,[28–31] a framework wherein emphasis is placed on the observed tendency for evolution to occur through changes in the timing of developmental patterns across generations. Evo-devo addresses developmental events from as early in life as possible, as well as organismal, cultural, and environmental influences on development at multiple levels throughout the lifespan.

The present approach to evaluating vocal development from the very beginning of life takes advantage of a technological innovation making possible all-day recordings of infant vocalizations and their families in the homes, a method that has been widely available for the past decade and a half[32,33] through the LENA Foundation https://web.archive.org/web/20230308172528/https://www.lena.org/. The LENA technology allows empirical evaluation of the amount of vocalization produced and heard by babies in maximally representative circumstances through automated categorization of vocalizations. A similar approach using all-day recording with adults has produced the tentative conclusion that there is little if any difference in the total amount of talk produced by men and women.[34] In both sexes, the rate recorded was enormous, ~16,000 words per day. Although infants in the first year rarely produce real words, we can estimate based on the human coding from CB-2020 that both baby boys and baby girls produce thousands of protophones daily, boys ~5–6 per minute, and girls ~4–5 per minute.

Volubility, which can be thought of as "talkativeness", in the term used by Leaper et al.,[18,19] is perhaps the most easily accessible measure affording comparison between boys and girls for a language-related

activity in the first two years. To the extent that infants may produce speech (real words), the utterances containing them will be included in the present paper within the volubility measure, which we shall term IVol (infant volubility or just volubility). The utterances of the first two years consist heavily of pre-speech vocalizations, called "protophones", defined first by exclusion, as referring to sounds other than speech, excluding the so-called "fixed signals"[35] such as crying or laughter, and also excluding vegetative sounds such as sneezing, burping, or hiccoughing.

The term "protophones" was coined[4] specifically to refer to all the presumable precursors to speech in infancy. Protophones begin in the first months of life with, for example, vowel-like sounds, squeals, growls, and raspberries; by the second half year, the protophones come also to include well-formed "canonical syllables" [ba], [aga], [dodo], [nene], and so on.[36,37] By referring to protophones in this technical sense, we avoid the ambiguity of the terms "babbling" or "speech-like vocalizations", which are often interpreted to refer only to utterances with canonical syllables, thus excluding the great majority of protophones that actually occur in the first year and that extend well into the second.[38]

Protophone volubility was the topic of CB-2020, based on human coding of longitudinal all-day recordings. The study reported that boys produced ~24% more protophones than girls across the first year, a difference that was not only highly statistically significant ($p < 0.001$) but also corresponded to a very large effect size ($d = 0.89$) when compared with the great majority of previously reported sex differences in language,[2,18,19] where even statistically significant differences have typically shown effect sizes <0.2. Since protophones change across the first year from being less like the sounds of speech in the first-half year to being more similar to speech during the canonical babbling stage of the second-half year, it was notable that CB-2020 found no difference between boys vs. girls in rates of canonical babbling. Thus, the results suggested infant males were more talkative, but not more advanced in the infraphonological content of their vocalizations once canonical babbling began.

The CB-2020 counts were based on human coding from ~1000 h of recorded material consisting of 5-min segments randomly selected from the larger entire longitudinal recording set (~6800 h). The data to be examined in the present work are based on LENA automated analysis of >450,000 h of recording, where every hour was automatically analyzed to yield estimated rates of infant vocalizations (see STAR Methods for description of the automated labeling procedure and relevant citations). Thus, the sample is >450 times larger than that of the CB-2020 report and covers semi-longitudinal data on 5899 infants from across the entire USA, nearly 60 times more infants than in the original study.

In addition to the advantage of an enormously larger body of data, the new study addresses recordings across two years rather than just the first year. The database supplies LENA automated estimates of 1) infant volubility (IVol), that is, the number of protophones as well as infant speech utterances judged by the algorithm to have occurred in each recording, a measure similar to that of CB-2020. The database also provides estimates of two measures not addressed in CB-2020: 2) conversational turns (CT), that is, the number of times an infant protophone or speech utterance was judged by the algorithm to have occurred within 5 s of an utterance produced by another speaker (usually a caregiver), and 3) adult word count (AWC), that is, the number of automatically estimated adult words spoken in the vicinity of the infant wearing the recorder.

In the data to be presented below, the terms "IVol" or "volubility" will be used always to encompass both protophones and early speech produced by infants—this lack of differentiation is necessary here, because LENA automated analysis cannot (yet) distinguish between protophones and speech produced by infants. CB-2020 also reported infant volubility in a way that included both protophones and infant speech. Similarly, the term "CT" in the present work will encompass events where infant protophones or speech alternated with utterances of other speakers.

The new data offer the occasion to delve further, and at much larger scale, into the interpretations of the apparently higher volubility of boys than girls in the first year, as reported in CB-2020, where three possible interpretations (there may be more) were considered: 1) because physical activity level is higher in boys than girls, greater volubility might be just another form of higher activity levels in boys[39,40]; 2) caregivers may talk more to boys than girls, eliciting higher volubility from boys; and 3) because boys are at higher risk of dying, especially in the first year,[41–43] boys may have been naturally selected across hominin history to produce vocal fitness signals at higher rates than girls, a compensation for their higher risk during that period.

The new data provide the opportunity to assess testable propositions that do not directly map onto the three interpretation options but that are expected to provide data relevant to their evaluation. The propositions, formulated to reflect likely outcomes based on the results from CB-2020 are.

1. boys will be found to be more voluble than girls in the first year of life;
2. more conversational turns will occur between boys and their caregivers than girls and their caregivers in the first year; and
3. higher AWC will tend to occur in recordings of boys than of girls in the first year.

CB-2020 offered no basis for predicting sex-differentiated outcomes for the second year.

RESULTS

Overview of the data for comparisons on sex

The present study evaluated >450,000 h of recording, the largest dataset we know of for a study of vocal or language development. The analysis produced the following numbers of machine-identified infant and adult utterances and interactive events: for IVol >56 million protophones and infant speech events, for CT >15 million turns, and for AWC nearly 600 million estimated adult words. These numbers of utterances, turns, and words are consistent with predictions that can be based on smaller prior studies using the LENA technology.[32,44]

Comparison of the IVol and CT data suggests that ~25% of protophones and infant speech were produced within conversational turns. This value should be interpreted cautiously because the definition of a conversational turn in the algorithm is based on alternations across time of adult and infant utterances, but the algorithm cannot determine whether adults and infants are actually conversing with each other—it can, however, reliably determine if the utterances are within 5 s of each other. Comparison of the CT data with the AWC data suggests that a CT occurred between an adult and an infant for about 2%–3% of adult words spoken in the vicinity of an infant. Again, since the algorithm cannot determine if parties are actually speaking to each other, the estimate of CTs needs to be interpreted cautiously. Comparison of the AWC with IVol data suggests that infants in the study heard more than 10 times as many words produced by adults as the utterances the infants themselves produced.

Age-related changes

Figures 1, 2, and 3 present the data month-by-month with breakdowns for male and female infants. Regarding Age, the IVol and CT variables (Figures 1 and 2) both showed significant increases from 2 to 24 months (rising ~77% for IVol, ~42% for CT), while the AWC variable (Figure 3) showed a decrease (falling ~19%). These patterns of rise or fall occurred for both male and female infants and for adults speaking in the presence of both male and female infants. Whereas increases in IVol and CT with age may seem plausible, one might question why the AWC would fall with age. One possibility is that very early in life, a caregiver needs to be close to the infant a great deal of the time, and thus caregiver voices are transmitted to the microphone of the recorder worn by the infant very often. As the infant becomes mobile and less in need of and less accessible to immediate caregiver attention, the number of adult utterances recorded may fall because adults are farther from the infant recorder, and signal-to-noise ratio of the adult voice may thus be reduced. Other investigators have also noticed a fall in AWC across time in their own LENA data; interestingly, they have also suggested that the amount of infant-directed speech appears to rise with age.[45]

As for the increase in CT across time as seen in Figure 2, the pattern may be partly the result of real increase in the amount of caregiver-infant vocal interaction with age. This idea is supported by the fact that the database of human coding associated with the 100 infants of CB-2020 suggests a notable increase in infant vocal turn-taking from the beginning until the end of the first year.

The first testable proposition on sex differences

Because IVol was the variable of greatest interest for comparison with CB-2020, we began by considering the pattern observable in the top panel of Figure 1, suggesting an interaction of Sex by Age. We proceeded then with a generalized estimating equations (GEE) analysis[46] on IVol, addressing Age as a continuous

CellPress
OPEN ACCESS
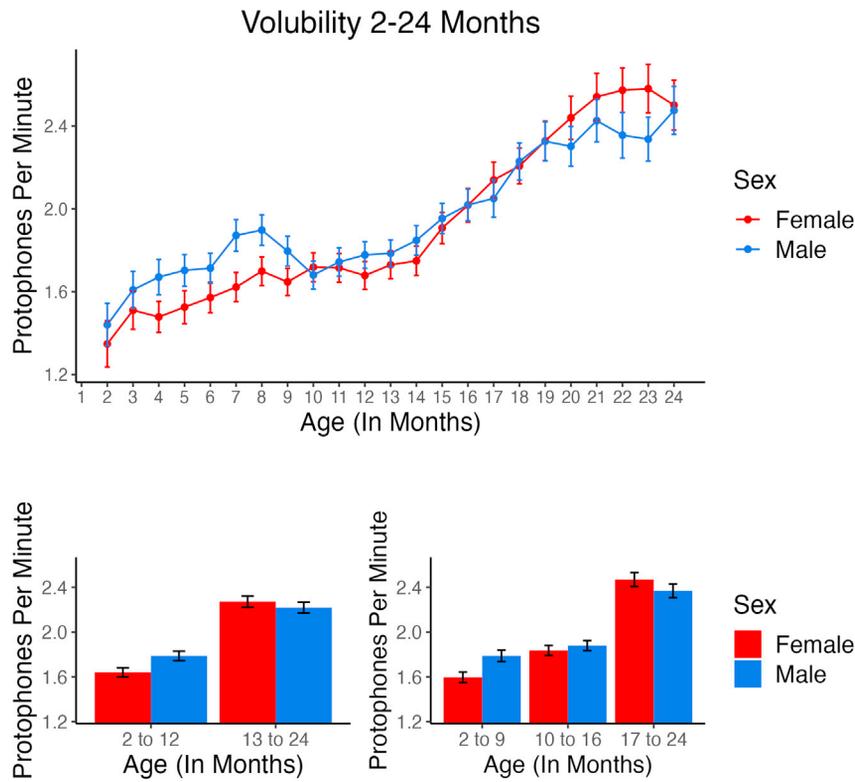
Figure 1. Volubility 2–24 months

In the upper panel, number of protophones per minute (Infant Volubility, IVol) is displayed month by month. IVol was found by the LENA algorithm to be higher for boys through about 9 months, but the pattern early in the second year favored the boys less, and later in the second year, it in fact revealed higher rates for girls. The rate of protophones per minute estimated by the algorithm increased across Age, with girls showing a greater increase than boys. Error bars represent 95% confidence intervals computed based on infant-level means for all available recordings at each Age. In the lower panels, mean values are displayed in the intervals that corresponded to the two GEE analyses, one for two Age intervals (left) and one for three (right). The means and CIs in the figure are not, however, the GEE estimated values, but were computed directly from the data at the infant level for all recordings available at each Age.

variable and Sex as a two-level fixed factor. It revealed, as expected based on Figure 1, a significant interaction of Age by Sex. The choice to treat Age as a variable with only two or three levels (see bottom panels of Figure 1) simplifies the presentation for all three testable propositions and makes hypothesis testing more straightforward than is possible when considering the data on all ages. For each of the three testable propositions, we present a primary GEE that analyzed a design with two fixed factors: Age (first year, 2–12 months vs. second year, 13–24 months) and Sex. For IVol, this analysis was selected because it corresponds to the most straightforward question raised by CB-2020 where boys showed higher volubility than girls across the first year. Data on the second year were not available in CB-2020, although it was available in the present data, so comparison across a first vs. second year split was deemed the most obvious choice for a simplified Age comparison. The secondary analysis, reported below for all three propositions, used three levels of Age (2–9 months, 10–16 months, and 17–24 months), a choice based on the rough appearance of the pattern of boy vs. girl outcomes on IVol across the entire Age range in Figure 1. A random factor in the designs for all three propositions was the individual infants, with all data for multiple recordings occurring for an individual within an age-month having been averaged across each age-month for the individual infants.

Figure 1 supplies detailed data for IVol across Age, computed at the infant level for each Age and supplying 95% confidence intervals as error bars. As suggested in the first testable proposition, boys did indeed produce more protophones/speech in the first year than girls, but in the second year the pattern was different. Toward the end of the second year, girls actually produced more protophones/speech than boys. The GEE analysis with two levels of Age, collapsing data across the first year and across the second, showed a
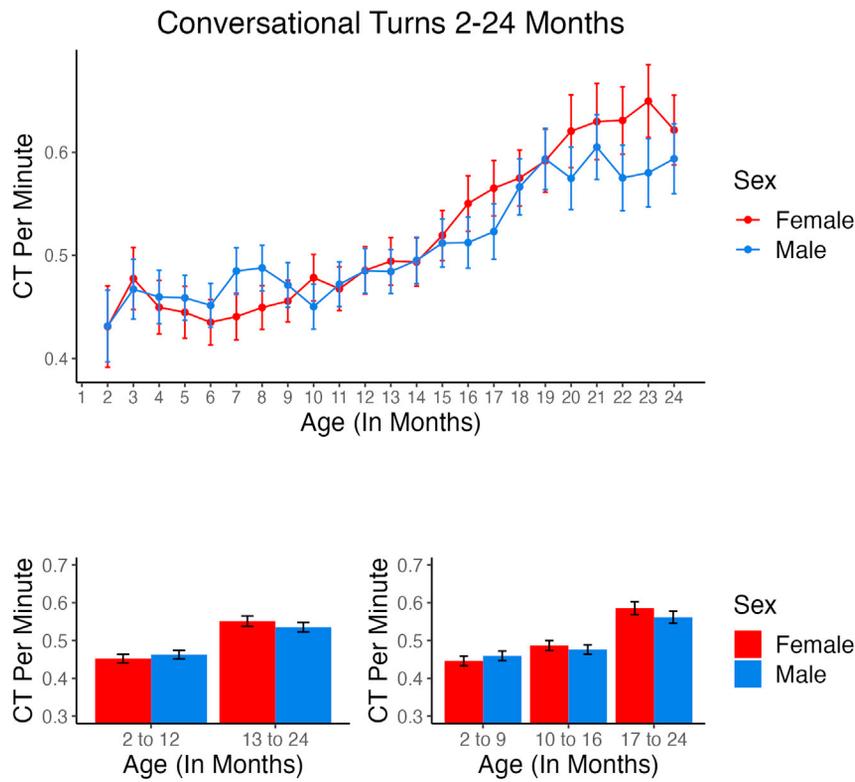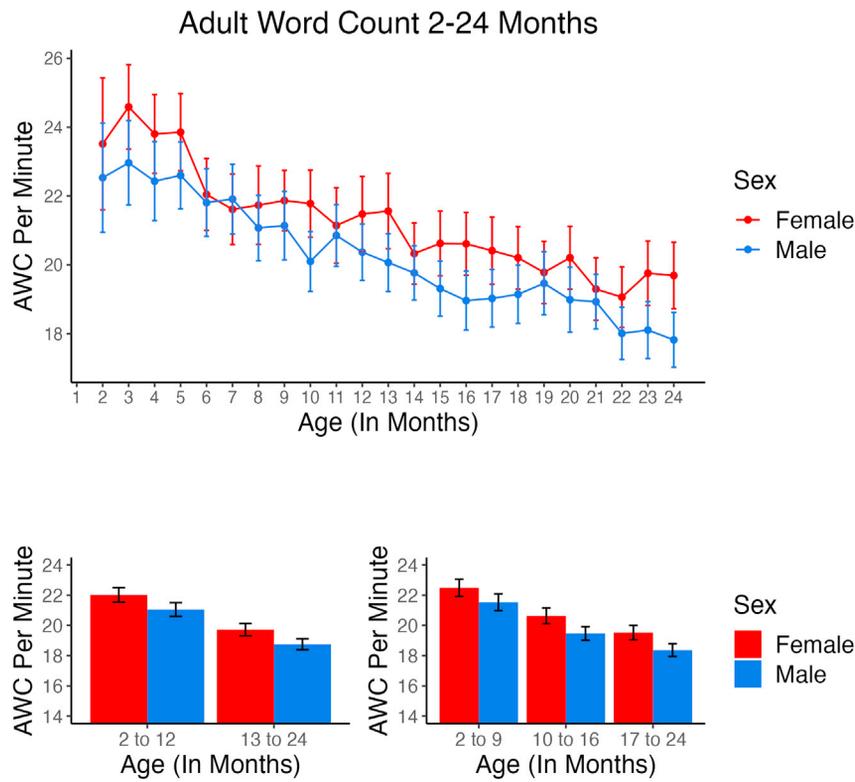
Figure 2. Conversational turn rate 2–24 months

In the upper panel, number of conversational turns per minute is displayed across all Age months. CT was found by the LENA algorithm to be higher for boys for some months of the first year, but the pattern thereafter tended to show higher rates for girls. As with the rate of protophones, conversational turns per minute estimated by the algorithm increased across Age, with girls showing a greater increase than boys, although the interaction of Sex and Age was not statistically significant. Only the Age effect was found to be significant. In the lower panels, mean values are displayed in the intervals that corresponded to the two GEE analyses, one for two Age intervals (left) and one for three (right). Error bars represent 95% confidence intervals computed based on infant-level means for all recordings at each Age.

statistically significant interaction ($p < 0.001$), reflected in the fact that boys were more voluble in the first year and girls more voluble in the second. With three levels of Age, collapsing data across 2–9 months, 10–16 months, and 17–24 months, the GEE also showed a statistically significant interaction ($p < 0.001$), reflected in the fact that boys were considerably more voluble in the first interval, and somewhat more voluble in the second, while the girls were more voluble as the end of the second year approached. Another way to interpret the interactions is to say that the girls showed increased volubility across Age to a greater extent than the boys.

For the two-Age-interval analysis, there was a significant Age effect ($p < 0.001$). For the three-Age-interval analyses, there was also a significant Age effect ($p < 0.001$), which applied to both the youngest group (2–9 months) vs. the middle group (10–16) and the youngest group (2–9 months) vs. the oldest (17–24). The Age effects reflect the considerable increase in the automated estimate of IVol across Age for both boys and girls. There was also a significant Sex effect for both the two-Age-interval and the three-Age-interval analyses ($p < 0.001$), indicating that boys overall showed higher volubility than girls, the interpretation of which needs to be mitigated by the girl's higher rates near the end of the second year.

The sizes of the main effects can be viewed both in terms of GEE estimates and in terms of computations from data at the infant level as manifest in Figure 1. For the two-Age-interval analysis, the GEE models estimated that the Ages differed by 0.34 protophones/speech per min while the Sexes differed by 0.12 protophones/speech per min. For the three-Age-interval analysis, GEE estimated a difference of 0.56 between the youngest group (2–9 months) and the oldest (17–24) and a difference of 0.09 between the youngest group (2–9 months) and the middle group (10–16). The estimated difference was 0.17 for the Sex effect.

CellPress
OPEN ACCESS

Figure 3. Adult word count 2–24 months

In the upper panel, number of adult words per minute spoken in the presence of male or female infants is displayed across all Ages. Number of estimated adult words per minute in the infant environment was found by the LENA algorithm to be higher for adults speaking in the presence of girls at 22 of 23 ages across the two years. In contrast to the rise in volubility and conversational turns across Age, adult word count estimated by the algorithm fell across Age, at similar rates for adults speaking in the presence of both boys and girls. The lower panels show the mean values corresponding to the Age intervals used in the GEE analyses. Error bars represent 95% confidence intervals computed based on infant-level means for all recordings at each Age.

However, to supply values that are more comparable to the effect sizes of prior sex and language research, we computed Cohen's d on various comparisons from the data as presented in Figure 1. These computations are presented in Table 1 under the heading IVol (Infant Volubility), along with the differences for mean protophones/speech per minute between the compared values. Thus, the Sex effect from 2 to 12 months was small ($d = 0.181$), favoring the boys, who produced 0.148 more protophones/speech per minute than the girls. Girls on the other hand showed higher volubility in the interval from 13 to 24 months ($d = -0.049$, 0.053 more protophones/speech per minute). Negative d's indicate the girls had higher values. The effect size for Age was much more notable than for Sex, with $d = 0.557$ for the analysis with two age intervals and $d = 0.707$ for the comparison between 2–9 months and 17–24 months for the analysis with three age intervals. Not displayed in Table 1 is perhaps the most noteworthy difference between males and females for the two-Age-interval analysis, that is, the change from the first (2–12 months) to the second (13–24 months) interval, where girls increased more than boys by 0.20 protophones/speech per minute ($d = 0.300$). Comparing the girls' greater IVol growth from the first interval of the three-age-interval analysis (2–9 months) to the last (17–24 months), we found an even greater difference of 0.29 protophones/speech per minute ($d = 0.408$).

The second testable proposition

Figure 2 suggests, consistent with our second proposition, that more conversational turns occurred between boys and caregivers than between girls and caregivers in the first year, but that the pattern, as with volubility, reversed in the second year. The differences were not as strong as in the case of volubility, and as a result, the interaction between Age and Sex was not statistically significant based on the GEE. The

iScience
Article

Table 1. Effect sizes

Comparisons	IVol (Infant Volubility)		CT (Conversational Turns)		AWC (Adult Word Count)	
	Cohen's d	Prot/min difference	Cohen's d	Turns/min difference	Cohen's d	Words/min difference
Two Age Intervals						
Sex, 2–12 months	0.181	0.148	0.047	0.010	−0.103	−0.964
Sex, 13–24 months	−0.049	−0.053	−0.056	−0.016	−0.116	−0.967
Age, 2–12 vs. 13–24	0.557	0.531	0.338	0.086	−0.260	−2.299
Three Age Intervals						
Sex, 2–9 months	0.229	0.191	0.063	0.014	−0.098	−0.945
Sex, 10–16 months	0.052	0.042	−0.046	−0.011	−0.133	−1.159
Sex, 17–24 months	−0.085	−0.101	−0.078	−0.024	−0.139	−1.157
Age, 2–9 vs. 10–16	0.200	0.165	0.127	0.029	−0.276	−2.539
Age, 10–16 vs. 17 to 24	0.551	0.561	0.341	0.092	−0.061	−0.522
Age, 2–9 vs. 17 to 24	0.707	0.726	0.455	0.121	−0.340	−3.061

Effect sizes and mean differences computed so that each infant's values for all recordings at a particular Age month were averaged before computing values across infants and Ages. Negative values indicate the girls had higher values.

only significant main effect was Age (p < 0.001, effect estimate = 0.049 turns/min for the two-age group analysis). The non-significant (by GEE) effect of change in female and male CTs across the two-Age-interval analysis showed girls outpacing boys in growth of CTs based on the data displayed in Figure 2 by 0.026 turns/min (d = 0.147), and for the three-Age-interval analysis, for the youngest age group vs. the oldest, by 0.038 turns/min (d = 0.204).

The third testable proposition

In Figure 3, the AWC results show that in general there was more adult talk in the vicinity of girls than of boys in both the first and the second years. There was no significant interaction of Age and Sex, but the GEE analysis revealed a statistically significant main effect of Sex (p < 0.02, effect estimate = 0.76 words/min with girls hearing more adult talk than boys) for the two-Age-interval analysis but not for the three-Age-interval analysis (p = 0.164, effect estimate = 0.51 words/min). Reanalyzing for main effects only (no interaction), the Sex effect was highly significant (p < 0.001) for both the two-Age-interval and three-Age-interval analyses (effect estimate = 0.96 and 0.94 words/min for the two analyses respectively). The Age effect in this re-analysis was highly significant for the two-Age-interval analyses (p < 0.001, effect estimate = 1.72 word/min) and also for the three-Age-interval analysis (p < 0.001, effect estimate = 2.00 for the youngest vs. the oldest group and p < 0.001, effect estimate = 1.13 for the youngest vs. the middle group), suggesting much more adult talk as estimated by the algorithm at the earlier ages than the later ones. Another way to assess the greater adult talk to girls is by noting that more than half of the 23 Ages shown in Figure 3 have error bars (95% confidence intervals) that do not overlap with means, in all cases showing more talk in the presence of girls than in the presence of boys. Moreover, the mean values were higher for girls at 22 of the 23 Ages.

DISCUSSION
Data summary

As predicted based on CB-2020, boys showed higher volubility (IVol ~10% higher) than girls in the first year. However, girls showed higher volubility in the second year (~7%). The contrasting patterns for the beginning and end of the period of observation corresponded to a statistically reliable interaction of Age and Sex. Similarly, the rate of CT was somewhat higher for boys in the first year and for girls in the second, though the corresponding interaction was not statistically reliable. On the other hand, the number of adult words (AWC) spoken in the vicinity of girls was higher across both the first and second years than in the case of boys. Comparing three Age categories (2–9 months, 10–16 months, and 17–24 months), we found the statistical contrast of boy vs. girl rates for IVol and CT in the youngest range vs. the oldest range was

stronger than in the case of the breakdown into just two Age ranges (2–12 and 13–24 months). The AWC difference favoring girls was, however, statistically strong for both the two-Age and the three-Age breakdowns. The month-by-month data presented in Figures 1, 2, and 3 suggest substantial variability from month-to-month for all the dependent variables and for the relations between the sexes.

Interpretations of sex differences in vocal development

The present data provide additional empirical perspective on the speculations from CB-2020 about the apparent early sex difference in volubility between boys and girls. One of those speculations was that boys might have been more voluble because male physical activity levels are higher than those of females in childhood.[40,47] In light of this fact, we might view vocal activity level as being just another type of physical activity level. However, the data from the second year of life appear to contradict that possible explanation because higher male volubility was not found beyond 16 months. On the contrary, female volubility was higher in the final months of the sampling. Furthermore, the greater physical activity level of boys does not diminish beyond the first year, but instead appears to be amplified as infancy and childhood progress.[40] The data do not thus offer support for a physical activity explanation for the sex difference in volubility.

Justification for the rejection of an activity-level explanation for higher male volubility in the first year is not airtight, however, first because the volubility data themselves, while statistically significant, were noisy across ages, and second because additional factors may come into play in the second year, potentially counterbalancing greater physical activity of boys. In particular, there may be greater social activity of girls talking in real words and sentences during the second year. Assuming it is true that girls acquire vocabulary faster than boys, presumably outdistancing them progressively in the second year,[17,48] then the availability of a larger lexicon in girls might inspire and elicit more interactive talk by girls with caregivers than by boys with caregivers in spite of any possible activity-level difference. This reasoning highlights the importance of evaluating not just volubility but also the extent of vocal interaction in both the first and second years.

The automated algorithms do not, unfortunately, offer an unambiguous evaluation of vocal interaction between caregiver and infant because automated analysis is unable at this stage of the technology's development to specifically designate caregiver speech directed to the infant wearing the recorder as opposed to speech directed to some other person. Even so, the CT measure provided by the LENA algorithm appears to provide a workable proxy for vocal interaction events between caregivers and infants. In general, the CT measure, in spite of its noisiness and its lack of specific designation of infant-directed speech, has yielded surprisingly powerful predictiveness for long-term language learning and cognitive development[49]; no other infant vocal measure has, to our knowledge, ever been reported to show such powerful predictiveness.

The results of the present work suggest a pattern with regard to sex and CT that roughly mirrors the pattern with regard to sex and volubility, though not as strongly. One possibility is that infant endogenous vocal tendencies across age, with boys more voluble in the first year and girls in the second, might elicit more vocal interactivity with caregivers for boys in the first year and girls in the second. Yet one might propose that the direction of the influence is the opposite; perhaps caregivers elicit more vocal interaction with boys in the first year and with girls in the second. Thus, the CT data do not clarify the direction of the observed pattern. The CT data, like the IVol data, also offer no support for the possibility that the higher vocal activity level of boys in the first year can be explained by higher physical activity levels in the boys, because boys show higher activity levels at all ages.[39,40]

On the other hand, the LENA algorithms do provide an interesting empirical comment, perhaps a refutation, of the idea that caregivers drive the pattern of CT, because the estimated number of adult words (AWC) heard by the infants was higher in general across both the first and the second years for *girls* than for boys. The fact that AWC does not distinguish between infant-directed speech and speech directed to others limits the generalizability of this refutation, but at present we have no better automated measure of likely caregiver vocal input to infants than AWC. From these data, then, we see no reason to assume that the male and female infant patterns of IVol and CT were primarily driven by caregiver elicitation.

The possibility that higher male volubility in the first year was driven by forces *within* the infant thus remains a viable interpretation of the evidence. Indeed, there is good reason to believe the human infant is not

primarily driven to vocalize by caregiver elicitation, nor by a desire to engage caregivers in vocal interchange.[50] Instead, the human infant appears to be, overwhelmingly, an endogenous vocalizer in the first year of life, a point to be explored in the next section.

The endogenous nature of infant vocalization and the Fitness Signaling Theory

Our currently favored explanation for the sex difference in volubility of infants (and the one favored in CB-2020) relies on reasoning about the high volubility of humans, both male and female, throughout life. It is important to inquire about the root of this tendency for humans to vocalize vastly more frequently and with far greater flexibility than our ape cousins.[26,51–56] Although the present paper does not supply evidence necessary to support the proposed answer to this inquiry, the currently favored explanation for the sex differences found here *depends* on that answer. Thus, it appears necessary to briefly explain the nature of the proposed reason for the existence of massive amounts of flexible human vocalization in order to formulate a workable explanation for the sex difference in early volubility.

Two of the present authors (Oller and Griebel) and John L. Locke have reasoned that when or sometime after ancient hominins first broke away from their ape cousins, a change occurred in vocal behavior of hominins whereby a tendency to vocalize more frequently and more freely emerged under natural selection pressure.[7,57] That greater tendency (to signal fitness) had to have had an advantage other than forming a foundation for language, because language did not yet exist. In accord with evo-devo reasoning,[31,58] it seems the change toward higher volubility must have occurred first in infants, who would then have grown up, to a greater extent in each generation, subject to the same selection pressure. Consequently, children, adolescents, and adults would have also shown a greater inclination to vocalize than their ape cousins, presumably supporting mate attraction, alliance formation, and other cooperative endeavors through vocal fitness signaling.

A key factor supporting the reasoning is that hominin infants were more altricial (helpless at birth) than the infants of their ape cousins because bipedalism had required narrowing of the hominin pelvis and had thus required a smaller head at birth, resulting in a necessarily slower developmental schedule, that is, a longer infancy and childhood.[59] Consequently, hominin infants were in greater need of long-term provisioning and care than their ape cousins. As a result, they were under increased selection pressure to signal their fitness through a variety of means, and vocalization, in particular the voluntary production of protophones, came to be one of the options for such fitness signaling.

One might ask why other apes, who were also somewhat altricial (but not as altricial as hominins), did not also respond to the pressure for fitness signaling by evolving voluntary vocalization. Indeed, the pressure to vocalize as a fitness signal would appear to apply to any species whose young are in need of provisioning by caregivers. But counter pressures must exist against vocalization, pressures against, for example, alerting predators by making too much noise or against having to develop the neurological foundations for voluntary vocalization. The pressure in favor of voluntary vocalization had to be sufficient in the hominins to outweigh such counter pressures, and the greater altriciality of the hominins tipped the balance, according to the Fitness Signaling Theory.

In any case, the reasoning behind the theory is fortified by two ecological circumstances under which ancient hominin infants were raised, circumstances that presumably increased further the relative advantage of vocal fitness signaling for hominin infants as opposed to the infants of other apes: 1) hominin groups are believed to have been larger than those of other apes across much of hominin history, affording more protection from predation[60]; and 2) cooperative breeding has been argued to have been more common in ancient hominins,[8,61] as it clearly is in modern humans, raising the premium on fitness signaling that could inform the many potential caregivers in a cooperative breeding environment of the wellness of infant vocalizers. In accord with the reasoning, hominin infant vocalizers competed *against each other* for caregiving investment. Favorable caregiver opinions about the viability of infants who supplied better fitness signals would have made those infants less likely to be neglected or abandoned than those whose vocal fitness signaling was less extensive or less effective.

This proposed inclination to produce protophones had to be "communicative" for the hominin infant in that potential caregivers had to hear the occurrence of those protophones, at least sometimes, and they had to interpret them as reliable indicators of infant wellness—indeed, they had to *be* reliable indicators

CellPress
OPEN ACCESS

of infant wellness in order that the tendency to produce them could have been stably evolved.[62] But at the same time, we do not propose that the infant had to *intend* to communicate with those vocalizations, at least not very frequently. The selection pressure was dependent on interpretations and subsequent actions of the caregivers who heard the vocalizations, even if infants did not normally intend them to be heard.

If the theory is on target, the pattern we propose to have occurred in ancient hominin infants may be present to an even greater degree in modern times. Modern human infants produce protophones at extremely high rates (~3500 per day, 4 to 5 per minute, every waking hour starting in the first month) that are bound to be noticed by caregivers often, whether the infant intends them to be heard or not.[24] Perhaps the most important surprise that has resulted from recent evaluation of protophone production in laboratory recordings is that infants produce ~70% of protophones without directing them to anybody.[50] In the cited investigation, even when caregivers attempted to elicit vocalizations after being instructed to do so by laboratory staff, infants produced the majority of their protophones (~60%) in such a way as to suggest they were not socially engaged with the caregiver.

Of course some protophones *are* produced interactively during periods when caregivers elicit them in face-to-face interaction,[63–65] but even in laboratory recordings where caregivers are always present, the rate of infant protophone production is at least as high when caregivers are silent and not interactive, as when caregivers attempt to elicit vocalization from their infants.[66] We have estimated, based on human coding of randomly selected segments from all-day audio recordings of 100 infants, that in fact caregivers engage in vocal interaction with infants only a very small proportion of the natural day at home[67]; caregivers spoke to infants often in 5-min segments when the infants were awake, but <20% of those segments contained any vocal responses from the infant, and <5% were deemed by the human coders to involve infant vocal responses during as much as half the time in those segments of observation. In contrast, >90% of the 5-min segments were deemed to involve vocal exploration, that is, protophone production by the infants in the absence of any sign of social directivity.[67]

These findings support the conclusion that the great majority of human infant protophones are produced endogenously. It thus seems likely that human infants have been naturally selected to treat protophone production as a kind of exploratory play, not unlike the playful activity of exploring objects with their hands, an activity that is shared with other primates,[68,69] although *vocal* exploratory play appears to be unique to humans among the apes. If, as we propose, the vocal exploratory play of human infants is inherently enjoyable, it is unnecessary for caregivers to elicit infant protophones, at least not often, in order to obtain significant fitness information from baby vocalizations.

How the Fitness Signaling Theory forms a possible basis for explanation of the sex difference in volubility of infants

The explanation we currently find most appealing for the greater volubility of baby boys is based on the supposition that boys are more vulnerable to death in the first year than girls, and consequently that it is to the advantage of boys to be especially active in producing protophone fitness signals in the first year. The higher volubility of boys is thus proposed to have been selected in order to secure caregiver investment that may help boys survive through that especially vulnerable period. Higher mortality in male infants is supported by a broad body of research,[6,42,43,70–72] but as usual with sex difference studies, there is variation in the degree of the differences reported across studies. While research suggests that girls may be generally more vulnerable (especially in some societies) to infanticide, the overall outcome in modern times appears to support the widely documented claim the boys are generally more vulnerable to death early in life.[73]

We evaluated data of the World Health Organization (WHO), currently covering the years from 1980 to 2017 at https://www.who.int/data/mortality/country-profile. The general pattern, supported by data compiled at the site from numerous nations on 6 continents, shows boys under the age of one year have notably higher death rates than girls. Of perhaps particular interest, data from Sub-Saharan Africa also show higher death rates in boys across the period from 1990 through 2020 (https://data.worldbank.org/indicator/SP.DYN.IMRT.FE.IN?locations=ZG).

According to the WHO data, the death rate for both sexes drops dramatically after the first year, and although boys continue to die at higher rates than girls in those subsequent years, the rates of death in

the first year are massively higher than in the remaining years of childhood. For example, in two nations (the USA and Brazil) with high populations and wide geographical distribution (and also differing substantially in GNP) available in the WHO database, we averaged across the years 2017, 2000, and 1980 for both nations, and found 27% more boys died in the first year of life than girls—this value should be adjusted downward somewhat for the 3% to 5% higher birth rate of boys.[74] Whereas boys continued in later years to die at higher rates than girls, the total number of deaths was >20 times higher in the first year than the combined deaths for the subsequent four years and >50 times higher than for the subsequent 10 years (individual year-by-year data are not provided at the WHO website beyond the first year so we can only present data clumped as indicated). Consequently, these modern data support the speculation that there should be higher selection pressure on boys than girls to produce signals of their wellness in the first year and that this special selection pressure on boys should be much higher in the first year than in subsequent years. The modern data on death rates of boys and girls cannot of course prove that boys died more than girls in the first year across hominin history. But data from archaeology are very spotty and hard to interpret,[75] yielding no evidence that relative death rates of boys and girls were different in our distant past from the way they are now.[76]

On fitness cues and fitness signals

We differentiate between cues and signals in keeping with the definitions of Maynard-Smith and Harper.[62] Cues are aspects or actions of any individual that may yield useful information about that individual to any observer. Cues are presumed not to require natural selection to make them constitute cues—but they may yield information about fitness anyway. Infant walking and crawling, for example, provide cues to wellness, but these actions were evolved as locomotion methods, not as wellness *signals*. Similarly, an infant's ability to sit up or manipulate objects by hand can provide cues to wellness. Infant cry, however, is presumably not merely a cue, because it was presumably naturally selected as a "signal" of infant state. It has in fact been argued that cry is a fitness signal, and to some extent it seems it must be.[77] Cry can be read by caregivers as a fitness signal to the extent that it occurs when needed, but if it occurs too much, it could be read as a negative indicator of wellness.

Similarly, we have proposed that protophones, naturally selected as fitness signals, must be judged as fitness signals by caregivers, who presumably deem infants to be well if they tend often to produce protophones in comfort and in pleasurable face-to-face interaction. Whiny protophones, like crying, are presumably judged by caregivers based on how they are produced—whining in discomfort may be judged positively, but an infant who whines too much may well be deemed less fit.

Evidence of robustness in protophone development

Extensive protophone production has been found in a vast array of studies of infants living in differing circumstances. For example, longitudinal comparison of English- and Spanish-learning infants as well as infants growing up bilingually revealed high volubility in all the groups,[78] and thus suggests high fitness signaling in all the groups. Similarly, infants in Korea and Taiwan show very similar patterns of protophone volubility to those in the USA.[79–81] Cross-cultural research evaluated face-to-face interactions with caregivers and their 5 and ½-month-old infants in 11 languages, finding that infants were very vocal in every language group.[82] While there are reports of very low infant-directed speech from caregivers in some societies,[82] we know of no data indicating very low infant volubility in any language or society. Infants of low socioeconomic status (SES) in the USA tend to produce lower volubility than infants of Mid SES, but the rates are relatively high in both cases.[38]

As for handicaps, infants show high levels of volubility at least in the first year for cases such as cleft palate,[83,84] Williams syndrome,[85] Fragile X syndrome,[86] autism,[87] and even bilateral profound deafness.[88] Infants born prematurely and still in neonatal intensive care begin protophone production at 2 months prior to due date (32 weeks gestational age), having been de-intubated shortly beforehand. Thus, they produce protophones at rates far higher than crying as soon as they can breathe on their own.[24] These indications of robustness of infant protophone production suggest deep foundations in the human species for vocal fitness signaling.

Special features of human vocal communication and babbling

Protophone production seems to be especially important in part, and perhaps primarily, because the functions of human language that are based on voluntary vocalization are extremely diverse. Although limited

iScience
Article

social functions beyond mate attraction and territorial defense do seem to be involved, although relatively rarely, in many cases of bird song[89] and other animal vocalization systems, human communication involves complexities that no other communication system we know of approaches. In modern humans, even the earliest protophones show massive functional flexibility, in that each of the vocal types (vowel-like sounds, squeals, and so on) can be produced in any emotional state. Consequently, each protophone type can transmit any kind of affect, ranging all the way from joy to fury.[90,91]

But notably, most of the time, human infants produce protophones in a state that appears to involve no deviation from emotional neutrality, expressing nothing more than interest in the exploration of the protophones themselves. That "interest," the apparent motivation to explore the sounds, may be the most important aspect of functional flexibility in human infant vocalization—it demonstrates the infant's vocal agency, that is, the infant's capacity and inclination to use vocalization voluntarily and not under the direction of any immediate need. If human infants did not have that capacity and inclination to vocalize freely, they would not have the capacity to learn language because every element of language—every syllable, word, and sentence—must be producible at any point in time, or else it would not be an element of language at all.[27] It seems reasonable to conclude that the babbled precursors to language, as well as all the later forms of mature linguistic expression, act as fitness signals by revealing states and capabilities to anyone listening. But language can transmit enormously more than fitness at the same time. Both male and female humans command a massive repertoire of vocally expressed illocutionary forces[92] as well as an indefinitely large repertoire of vocal semantic expressions that can be used to support cooperative foraging, hunting, cooking, building, warring, and so on.

Why should there be any sexual dimorphism in human protophones given the great similarity of language capability in the mature of both sexes? In accord with our reasoning above, the differences may be local to a narrow age range within the first year, and they may result from differences in health vulnerability of males and females resulting in stronger selection pressure on boys to signal their fitness. These differences do not result in obvious *communicative* limitations on females at any point. Indeed, if anything, it appears females may still have the edge overall in language competition between the sexes.

Differences between outcomes from LENA Start and LENA Home

The data for the present work were derived from recordings made during two programs of intervention conducted by the LENA Foundation, LENA Start and LENA Home (see STAR Methods for descriptions). The two programs produced somewhat different patterns of results: LENA Start showed 10%, 19%, and 6% higher than LENA Home values for IVol, CT, and AWC, respectively. Furthermore, the sex differences for both IVol and CT were stronger for LENA Start than LENA Home, and the pattern for AWC where girls heard more adult words was stronger for LENA Home. There are many possible reasons for such differences, although it is impossible to tie them down given that we have so little information on possible differing demographic factors and/or differing program implementations. A tentatively important group difference is based on a proxy for SES, the ADI (Area Deprivation Index).[93] Nearly 32% of caregivers reported ADI, yielding an average at the 38th percentile nationally (LENA Start infants, 46th percentile; LENA Home infants 26th). Given this scanty and weak information and the fact that it was not possible to ensure that the presumed intervention protocols were implemented rigorously and in accord with protocols across the many sites of intervention all over the USA, we offer no explanation for the different patterns across the two sources of data. Instead, we caution that a wide variety of social factors may modulate the sex differences reported here.

Differences in volubility as a function of sex in the present paper compared to differences found in CB-2020

In the current data, boys produced 9% more protophones than girls across the first year (the IVol variable), while in CB-2020, the difference favoring boys was 24%. The effect sizes as reflected in Cohen's *d* were also much higher in CB-2020 for the same age range ($d = 0.18$ for the current data vs. $d = 0.89$ for CB-2020). The effect size shown in the current data is comparable to that found in the bulk of prior studies on sex differences in language across the lifespan, while the CB-2020 data showed an effect size several times larger.

Perhaps the difference is due to the higher sensitivity of human coding as conducted for CB-2020 compared with the automated method that provided the current data. Human listeners often recognize who produces utterances that are overlapped in the acoustic signal, while the automated method does

not have that capability. As a consequence, the human coded data may more accurately reflect actual rates. The difference may also be due to differences in demographics of the samples. For example, although there is no unambiguous way to compare SES across the samples, it seems clear that the CB-2020 data were based on recordings from relatively high SES (maternal education levels were clearly above the national average), while the data from the present work were based on relatively low SES. The suggestion that SES might affect sex differences in volubility is of course speculative, but other data support the idea that infants from low SES families show lower volubility.[44,94,95] Perhaps low volubility in a group tends to minimize volubility differences between subgroups.

Differences in volubility of all infants (regardless of sex) and volubility as a function of age in the present paper compared to differences found in CB-2020

The IVol value as estimated by the LENA algorithm in the present work (∼1.4–2.4 per min) was lower by more than a factor of two than that found in studies of infants in the first year of life using human coding of randomly selected segments from LENA recordings.[1,24,26] The differences have at least two major sources: 1) human listeners can often differentiate infant voices from other voices or sounds superimposed upon them, while the automated algorithm often cannot differentiate them, producing an overlap categorization (see STAR Methods), and thus the number of infant utterances identified by the algorithm is necessarily reduced; and 2) the cited human-coding studies excluded all segments where the infant was deemed by the coder to be asleep (∼20%–30% of segments) and where mean volubility was only ∼2% of the mean volubility for wakeful segments. In contrast, the automated algorithm does not eliminate sleep segments, yielding lower volubility per unit time (presumably again by ∼20%–30%). There may be additional so far unidentified reasons for the differences in IVol based on the LENA algorithm and human coding.

Another notable difference between the human-coding results and those of the automated algorithm is that no increase in infant volubility was discerned across the first year in the human coding of CB-2020 nor in other human-coding studies with LENA recordings.[24,96] It may be possible to explain at least part of this difference between human-coding studies and the present results where IVol increased with age based on the fact that sleep segments are more common at younger ages than older ones. For example, at 0 months ∼31% of 5-min segments were judged to involve an infant asleep, compared to ∼23% of segments at 12 months based on data from above cited human-coding studies of infants recorded all-day approximately monthly across the first year.[96] Sleep segments are excluded in estimates supplied by human-coding studies such as those cited, whereas sleep segments were not identified and thus could not be excluded in the computations based on the automated algorithm used for the present study. There may be additional reasons for the increase in IVol observed in the present study as opposed to the relatively constant IVol across age in the first year in studies using human coding. As for the second year of life, the current data show substantial increases in IVol, but extensive human coding data are as yet unavailable for comparison.

Conclusions

The present evidence supports the findings reported in CB-2020 based on human coding of all-day recordings by supplying evidence of a sex difference in volubility of human infants. Boys produced higher volubility in the first year, whereas by the end of the second year, girls showed higher volubility. In seeking a possible reason to explain the boys' higher volubility in the first year, the evidence argues against the possibility that higher physical activity of boys could account for the difference; boys show higher physical activity levels throughout childhood, but the evidence does not suggest higher volubility of boys except in the first year and perhaps a few months beyond that. Furthermore, the evidence presented does not support the supposition that caregivers elicit more vocalization from boys in the first months of life because the automated analyses suggest adults spoke more often in the presence of girls than boys across both of the first two years.

A third possibility to explain higher volubility of boys early in life is suggested by the facts that boys are more vulnerable to death early in life and that mortality is much higher in the first year than in subsequent years. Thus, we have reasoned that boys may have been naturally selected to produce higher volubility in the first months of life than girls as a signal of fitness, presumably eliciting additional caregiver investment and thus partially offsetting male vulnerability to dying in the first year. This interpretation, while preliminary, is consistent with an evo-devo interpretation of human vocal evolution and development, where the massive amounts of vocalization in both modern and presumably ancient humans are seen as having

CellPress
OPEN ACCESS

been selected in the distant past as fitness signals. Our reasoning suggests that by establishing a firm foundation of voluntary vocal activity used to signal fitness, hominins laid the groundwork for additional developments that ultimately led to language. Similarly, we reason that modern human infants lay the groundwork for their own subsequent development of language also by their inclination and capacity to explore vocalization voluntarily from as soon as they can breathe independently. We propose that the apparent existence of early sex differences in vocal development calls for invoking potentially differential requirements for survival of male and female infants in the first year of life.

Limitations of the study

The present work was embedded in intervention efforts that allowed access only to limited information about the participating families. Importantly, the all-day recordings were erased after automated analysis because we did not have permission to keep them. While we had access to infant sex and infant age at the time of each recording, we had no access to data on a variety of factors (number, birth order, and sex of other children in the family, number, sex, and ages of caregivers, and so on) that might have been useful in helping with interpretation of the outcomes. Because the recordings were not available for human coding, we were required to work with the results of the automated analysis exclusively. Also, there was no opportunity for conducting the study with random assignment of infants to an experimental vs. a control group.

The present work is based entirely on automated analyses that are modeled on human coding and thus are designed to simulate results that can be obtained with human coding. Further research based on human coding of large samples to assess possible sex differences in vocal development is surely desirable. Of course there is a trade-off: one can analyze huge amounts of recorded data with automated methods, but one can obtain more reliable data with human coding of smaller amounts of data. Converging evidence from both methods is thus desirable.

The automated method is constantly in development and will continue to improve into the future. Among the possibilities for improvement is expansion of the number of variables that can be categorized and validated for performance with respect to human coding. Of particular interest will be validation of automated categorization for: 1) infant vocalizations, conversational turns, and adult utterances that occur during overlap of voices, 2) infant crying and laughter as well as vegetative vocalizations, 3) affectively negative protophones (whining), and 4) caregiver utterances that are specifically directed to the infant wearing the recorder. None of these can be currently assessed with LENA automated categorization with a sufficient level of confidence to justify analyses such as the ones delineated in the present article for IVol, CT, and AWC.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Sample size estimation
  - How subjects/samples were allocated to experimental groups
  - General information on participants
- METHOD DETAILS
  - **Numbers and lengths of** recordings
  - Recording device and automated analyses
  - **No original recordings** available
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106884.

CellPress
OPEN ACCESS

iScience
Article

DECLARATION OF INTERESTS

S.H. is President and CEO of the LENA Foundation. J.G. is Chief Research and Evaluation Officer of the LENA Foundation. J.A.R. is Chief Statistician of the LENA Foundation. D.K.O. and S.F.W. are unpaid members of the Scientific Advisory Board of the LENA Foundation. S.F.W. is also an unpaid member of the LENA Foundation Board. The LENA Foundation holds patents on the LENA Digital Language Processor (DLP, the recorder) and the software that is used to automatically categorize utterances from recordings made with the DLP. More information can be found at the following link: https://www.lena.org/patents/.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Oller, D.K., Griebel, U., Bowman, D.D., Bene, E., Long, H.L., Yoo, H., and Ramsay, G. (2020). Infant boys found to be more vocal than infant girls. Curr. Biol. *30*, 426–427. https://www.cell.com/current-biology/fulltext/S0960-9822(20)30419-X.

2. Hyde, J.S., and Linn, M.C. (1988). Gender differences in verbal ability: a meta-analysis. Psychol. Bull. *104*, 53–69. https://doi.org/10.1037/0033-2909.104.1.53.

3. Locke, J.L. (1993). The Child's Path to Spoken Language, First Edition (Harvard University Press). https://www.hup.harvard.edu/catalog.php?isbn=9780674116399.

4. Oller, D.K. (2000). The Emergence of the Speech Capacity (originally Lawrence Erlbaum Associates, now Routledge, Taylor and Francis Group).

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

6. Naeye, R.L., Burt, L.S., Wright, D.L., Blanc, W.A., and Tatter, D. (1971). Neonatal mortality, the male disadvantage. Pediatrics *48*, 902–906. https://doi.org/10.1542/peds.48.6.902.

7. Locke, J.L. (2006). Parental selection of vocal behavior: crying, cooing, babbling, and the evolution of language. Hum. Nat. *17*, 155–168. https://link.springer.com/article/10.1007/s12110-006-1015-x.

8. Oller, D.K., and Griebel, U. (2021). Functionally flexible signaling and the origin of language. Front. Psychol. *11*, 1–10. https://doi.org/10.3389/fpsyg.2020.626138.

9. Fisher, M.S. (1934). Language Patterns of Preschool Children (Monograph Society for Research in Child Development). https://www.jstor.org/stable/20150257.

10. McCarthy, D. (1953). Some possible explanations of sex differences in language development and Disorders. J. Psychol. *35*, 155–160. https://doi.org/10.1080/00223980.1953.9712848.

11. Anastasi, A. (1958). Differential Psychology, 3 Edition. https://www.abebooks.com/book-search/title/differential-psychology/author/anastasi-a/.

12. Bayley, N., and Oden, G.C. (1955). The maintenance of intellectual ability in gifted adults. J. Gerontol. *10*, 91–107. https://doi.org/10.1093/geronj/10.1.91.

13. Ramist, L., and Arbeiter, S. (1986). Profiles, College-Bound Seniors, 1985 (College Entrance Examination Board). https://eric.ed.gov/?id=ED282474.

14. Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: a critical review. Brain Lang. *108*, 175–183. https://doi.org/10.1016/j.bandl.2008.07.001.

15. Galsworthy, M.J., Dionne, G., Dale, P.S., and Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. Dev. Sci. *3*, 206–215. https://doi.org/10.1111/1467-7687.00114.

16. Bornstein, M.H., Hahn, C.-S., and Haynes, O.M. (2004). Specific and general language performance across early childhood: stability and gender considerations. First Lang. *24*, 267–304. https://doi.org/10.1177/0142723704045681.

17. Eriksson, M., Marschik, P.B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., and Gallego, C. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. Br. J. Dev. Psychol. *30*, 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x.

18. Leaper, C., and Ayres, M.M. (2007). A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech. Pers. Soc. Psychol. Rev. *11*, 328–363. https://doi.org/10.1177/1088868307302221.

19. Leaper, C., and Smith, T.E. (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. Dev. Psychol. *40*, 993–1027. https://doi.org/10.1037/0012-1649.40.6.993.

20. Cameron, D. (2009). Sex/gender, language and the new biologism. Appl. Linguist. *31*, 173–192. https://doi.org/10.1093/applin/amp022.

21. Schudson, Z.C., Beischel, W.J., and van Anders, S.M. (2019). Individual variation in gender/sex category definitions. Psychology of Sexual Orientation and Gender Diversity 6, 448–460. https://doi.org/10.1037/sgd0000346.

22. Locke, J.L. (2011). Duels and Duets: Why Men and Women Talk So Differently (Cambridge University Press). https://www.cambridge.org/core/books/duels-and-duets/8F31691CCE8D633D2315E23D6E605644.

23. Brescoll, V.L. (2011). Who takes the floor and why: gender, power, and volubility in organizations. Adm. Sci. Q. 56, 622–641. https://doi.org/10.1177/0001839212439994.

24. 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. https://doi.org/10.1038/s41598-019-51352-0.

25. Griebel, U., Pepperberg, I.M., and Oller, D.K. (2016). Developmental plasticity and language: a comparative perspective. Top. Cogn. Sci. 8, 435–445. https://doi.org/10.1111/tops.12200.

26. 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. https://doi.org/10.3389/fpsyg.2019.00729.

27. Oller, D.K., Griebel, U., and Warlaumont, A.S. (2016). In 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, 8, W.D. Gray, S. Issue, D. Kimbrough Oller, R. Dale, and U. Griebel, eds., pp. 382–392. https://doi.org/10.1111/tops.12200.

28. Carroll, S.B. (2005). Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom (W. W. Norton)). https://wwnorton.com/books/Endless-Forms-Most-Beautiful/.

29. Gottlieb, G. (2002). Developmental–behavioral initiation of evolutionary change. Psychol. Rev. 109, 211–218. https://pubmed.ncbi.nlm.nih.gov/11990317/.

30. Newman, S.A., and Müller, G.B. (2000). Epigenetic mechanisms of character origination. J. Exp. Zool. 288, 304–317. https://doi.org/10.1002/1097-010X(20001215)288:4%3C304::AID-JEZ3%3E3.0.CO.

31. Locke, J.L. (2009). Evolutionary developmental linguistics: naturalization of the faculty of language. Lang. Sci. 31, 33–59. https://www.sciencedirect.com/science/article/pii/S0388000107001039.

32. Gilkerson, J., and Richards, J.A. (2007). The Infoture Natural Language Study. Infoture Technical Reports, ITR-02-1. ITR-02-1.

https://web.archive.org/web/20230317215404/https://www.lena.org/technology/#tech-reports.

33. Zimmerman, F.J., Gilkerson, J., Richards, J.A., Christakis, D.A., Xu, D., Gray, S., and Yapanel, U. (2009). Teaching by listening: the importance of adult-child conversations to language development. Pediatrics 124, 342–349. https://doi.org/10.1542/peds.2008-2267.

34. Mehl, M.R., Vazire, S., Ramírez-Esparza, N., Slatcher, R.B., and Pennebaker, J.W. (2007). Are women really more talkative than men? Science 317, 82. https://doi.org/10.1126/science.1139940.

35. Tinbergen, N. (1952). Derived activities: their causation, biological significance, origin and emancipation during evolution. Q. Rev. Biol. 27, 1–32. https://www.jstor.org/stable/2812621.

36. Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In Child Phonology, 1, G. Yeni-Komshian, J. Kavanagh, and C. Ferguson, eds. (Academic Press), pp. 93–112. https://www.sciencedirect.com/science/article/abs/pii/B9780127706016500115.

37. Stark, R.E. (1980). Stages of speech development in the first year of life. In Child Phonology, 1, G. Yeni-Komshian, J. Kavanagh, and C. Ferguson, eds. (Academic Press), pp. 73–90.

38. 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 Behav. Dev. 16, 297–315. https://www.sciencedirect.com/science/article/pii/0163638393800379?via%3Dihub.

39. Campbell, D.W., and Eaton, W.O. (1999). Sex differences in the activity level of infants. Infant Child Dev. 8, 1–17. https://doi.org/10.1002/(sici)1522-7219(199903)8:1<1::aid-icd186>3.0.co;2-o.

40. Eaton, W.O., and Enns, L.R. (1986). Sex differences in human motor activity level. Psychol. Bull. 100, 19–28. https://doi.org/10.1037/0033-2909.100.1.19.

41. Hammoud, E.I. (1965). Studies in fetal and infant mortality: I. A methodological approach to the definition of perinatal mortality. Am. J. Public Health Nation's Health 55, 1012–1023. https://doi.org/10.2105/ajph.55.7.1012.

42. Drevenstedt, G.L., Crimmins, E.M., Vasunilashorn, S., and Finch, C.E. (2008). The rise and fall of excess male infant mortality. Proc. Natl. Acad. Sci. USA 105, 5016–5021. https://doi.org/10.1073/pnas.0800221105.

43. Zhao, D., Zou, L., Lei, X., and Zhang, Y. (2017). Gender differences in infant mortality and neonatal morbidity in mixed-gender twins. Sci. Rep. 7, 8736. https://doi.org/10.1038/s41598-017-08951-6.

44. Gilkerson, J., Richards, J.A., Warren, S.F., Montgomery, J.K., Greenwood, C.R., Kimbrough Oller, D., Hansen, J.H.L., Paul,

T.D., and Paul, T.D. (2017). Mapping the early language environment using all-day recordings and automated analysis. Am. J. Speech Lang. Pathol 26, 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169.

45. Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A.S., and Amatuni, A. (2019). What do north American babies hear? A large-scale cross-corpus analysis. Dev. Sci. 22, e12724. https://doi.org/10.1111/desc.12724.

46. Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22. https://biostat.jhsph.edu/~jleek/teaching/2011/754/reading/liangandzeger.pdf.

47. Else-Quest, N.M., Hyde, J.S., Goldsmith, H.H., and Van Hulle, C.A. (2006). Gender differences in temperament: a meta-analysis. Psychol. Bull. 132, 33–72. https://doi.org/10.1037/0033-2909.132.1.33.

48. Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L.V. (2010). Sources of variability in children's language growth. Cognit. Psychol. 61, 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002.

49. Gilkerson, J., Richards, J.A., Warren, S.F., Oller, D.K., Russo, R., and Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. Pediatrics 142, e20174276. https://doi.org/10.1542/peds.2017-4276.

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

51. Hauser, M. (1996).). The evolution of Communication (MIT). https://mitpress.mit.edu/9780262581554/the-evolution-of-communication/.

52. Crockford, C., and Boesch, C. (2003). Context-specific calls in wild chimpanzees, Pan troglodytes verus: analysis of barks. Anim. Behav. 66, 115–125. https://doi.org/10.1006/anbe.2003.2166.

53. De Waal, F.B. (1988). The communicative repertoire of captive bonobos (Pan paniscus), compared to that of chimpanzees. Beyond Behav. 106, 183–251. https://doi.org/10.1163/156853988X00269.

54. Laporte, M.N.C., and Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. Dev. Sci. 14, 1220–1234. https://doi.org/10.1111/j.1467-7687.2011.01069.x.

55. Stewart, K.J., and Harcourt, A.H. (1994). Gorilla's vocalizations during rest periods: signals of impending departure? Beyond Behav. 130, 29–40. https://www.jstor.org/stable/4535204.

56. Tagliatelatela, J.P., Reamer, L., Schapiro, S.J., and Hopkins, W.D. (2012). Social learning of a communicative signal in captive chimpanzees. Biol. Lett. 8, 498–501. https://doi.org/10.1098/rsbl.2012.0113.

57. 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. (Sage Publications), pp. 135–166. https://sk.sagepub.com/books/evolutionary-perspectives-on-human-development-2e/n5.xml.

58. 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, 2056–2068. https://doi.org/10.1098/rstb.2011.0035.

59. Locke, J.L., and Bogin, B. (2006). Language and life history: a new perspective on the evolution and development of linguistic communication. Behav. Brain Sci. 29, 259–280. https://doi.org/10.1017/S0140525X0600906X.

60. Dunbar, R. (1996). Grooming, Gossip, and the Evolution of Language (Harvard University Press). https://www.hup.harvard.edu/catalog.php?isbn=9780674363366.

61. Hrdy, S.B. (2016). Variable postpartum responsiveness among humans and other primates with "cooperative breeding": a comparative and evolutionary perspective. Horm. Behav. 77, 272–283. https://doi.org/10.1016/j.yhbeh.2015.10.016.

62. Maynard Smith, J., and Harper, D. (2003). Animal Signals (Oxford University Press).

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

64. Stern, D.N. (1974). Mother and infant at play: the dyadic interaction involving facial, vocal, and gaze behaviors. In The effect of the infant on its caregiver, M. Lewis and L.A. Rosenblum, eds. (Wiley), pp. 187–213.

65. Trevarthen, C. (1974). Conversations with a two-month-old. New Scientist 2, 230–235.

66. Iyer, S.N., Denson, H., Lazar, N., and Oller, D.K. (2016). Volubility of the Human Infant: Effects of Parental Interaction (Or Lack of it) (Clinical Linguistic & Phonetics), pp. 1–19. https://doi.org/10.3109/02699206.2016.1147082.

67. Long, H.L., Ramsay, G., Griebel, U., Bene, E.R., Bowman, D.D., Burkhardt-Reed, M.M., and Oller, D.K. (2022). Perspectives on the origin of language: infants vocalize most during independent vocal play but produce their most speech-like vocalizations during turn taking. PLoS One 17, e0279395. https://doi.org/10.1371/journal.pone.0279395.

68. Cenni, C., Casarrubea, M., Gunst, N., Vasey, P.L., Pellis, S.M., Wandia, I.N., and Leca, J.-B. (2020). Inferring functional patterns of tool use behavior from the temporal structure of object play sequences in a non-human primate species. Physiol. Behav. 222, 112938. https://doi.org/10.1016/j.physbeh.2020.112938.

69. Lewis, K.P. (2000). A comparative study of primate play behaviour: implications for the study of cognition. Folia Primatol. 71, 417–421. https://doi.org/10.1159/000052740.

70. Khoury, M.J., Marks, J.S., McCarthy, B.J., and Zaro, S.M. (1985). Factors affecting the sex differential in neonatal mortality: the role of respiratory distress syndrome. Am. J. Obstet. Gynecol. 151, 777–782. https://doi.org/10.1016/0002-9378(85)90518-6.

71. Mage, D.T., and Donner, M. (2004). The X-linkage hypotheses for SIDS and the male excess in infant mortality. Med. Hypotheses 62, 564–567. https://doi.org/10.1016/j.mehy.2003.10.018.

72. Tashiro, A., Yoshida, H., and Okamoto, E. (2019). Infant, neonatal, and postneonatal mortality trends in a disaster region and in Japan, 2002–2012: a multi-attribute compositional study. BMC Publ. Health 19, 1085. https://doi.org/10.1186/s12889-019-7443-4.

73. Fuse, K., and Crenshaw, E.M. (2006). Gender imbalance in infant mortality: a cross-national study of social structure and female infanticide. Soc. Sci. Med. 62, 360–374. https://doi.org/10.1016/j.socscimed.2005.06.006.

74. Chao, F., Gerland, P., Cook, A.R., and Alkema, L. (2019). Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. Proc. Natl. Acad. Sci. USA 116, 9303–9311. https://doi.org/10.1073/pnas.1812593116.

75. Mays, S., and Faerman, M. (2001). Sex identification in some putative infanticide victims from roman britain using ancient DNA. J. Archaeol. Sci. 28, 555–559. https://doi.org/10.1006/jasc.2001.0616.

76. Scott, E. (1999). The Archaeology of Infancy and Infant Death (Archaeopress, Publishers of British Archaeological Reports). http://primo.getty.edu/GRI:GETTY_ALMA21114864360001551.

77. Godfray, H.C.J. (1991). Signalling of need by offspring to their parents. Nature 352, 328–330. https://doi.org/10.1038/352328a0.

78. Oller, D.K., and Eilers, R.E. (1982). Similarity of babbling in Spanish- and English-learning babies. J. Child Lang. 9, 565–577. https://doi.org/10.1017/S0305000900004918.

79. Ha, S., and Oller, D.K. (2019). Canonical babbling in Korean-acquiring infants at 4-9 Months of age. Commun. Sci. Disord. 24, 1–8. https://doi.org/10.12963/csd.19577.

80. Lee, C.-C., Jhang, Y., Relyea, G., Chen, L.-M., and Oller, D.K. (2018). Babbling development as seen in canonical babbling ratios: a naturalistic evaluation of all-day recordings. Infant Behav. Dev. 50, 140–153. https://doi.org/10.1080/15475441.2016.1180983 NIHMSID #821602.

81. Lee, C.-C., Jhang, Y., Chen, L.M., 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. Lang. Learn. Dev. 13, 100–126. https://doi.org/10.1080/15475441.2016.1180983.

82. Bornstein, M.H., Putnick, D.L., Cote, L.R., Haynes, O.M., and Suwalsky, J.T.D. (2015). Mother-infant contingent vocalizations in 11 countries. Psychol. Sci. 26, 1272–1284. https://doi.org/10.1177/0956797615586796.

83. Chapman, K.L., Hardin-Jones, M., Schulte, J., and Halter, K.A. (2001). Vocal development of 9-month-old babies with cleft palate. J. Speech Lang. Hear. Res. 44, 1268–1283. https://doi.org/10.1044/1092-4388(2001/099.

84. Ha, S., and Oller, D.K. (2021). Longitudinal study of vocal development and language environments in infants with cleft palate. Cleft Palate-Craniofacial J. 1055665621042513. https://doi.org/10.1177/10556656211042513.

85. 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. Dev. Sci. 4, 158–164. https://doi.org/10.1111/1467-7687.00161.

86. Belardi, K., Watson, L.R., Faldowski, R.A., Hazlett, H., Crais, E., Baranek, G.T., McComish, C., Patten, E., and Oller, D.K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with fragile X syndrome at 9 -12 Months of age. J. Autism Dev. Disord. 47, 1193–1206. https://doi.org/10.1007/s10803-017-3033-4.

87. 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. J. Autism Dev. Disord. 44, 2413–2428. https://doi.org/10.1007/s10803-014-2047-4.

88. 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. https://doi.org/10.17955/TVR.108.2.603.

89. Kaplan, G. (2018). Development of meaningful vocal signals in a juvenile territorial songbird (Gymnorhina tibicen) and the dilemma of vocal taboos concerning neighbours and strangers. Animals 8, 228. https://doi.org/10.3390/ani8120228.

90. Jhang, Y., and Oller, D.K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. Front. Psychol. 8, 300. https://doi.org/10.3389/fpsyg.2017.00300.

91. 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. Proc. Natl. Acad. Sci. USA 110,

6318–6323. https://doi.org/10.1073/pnas.300337110.

92. Austin, J.L. (1962). How to Do Things with Words (Oxford Univ. Press). https://pure.mpg.de/rest/items/item_2271128/component/file_2271430/content.

93. Maroko, A.R., Doan, T.M., Arno, P.S., Hubel, M., Yi, S., and Viola, D. (2016). Integrating social determinants of health with treatment and prevention: a new tool to assess local area deprivation. Prev. Chronic Dis. 13, 160221. https://doi.org/10.5888/pcd13.160221external icon.

94. Hart, B., and Risley, T.R. (1995). Meaningful Differences in the Everyday Experience of Young American Children (Paul H. Brookes)).

95. Oller, D.K., Eilers, R.E., Steffens, M.L., Lynch, M.P., and Urbano, R. (1994). Speech-like vocalizations in infancy: an evaluation of potential risk factors. J. Child Lang. 21, 33–58. https://doi.org/10.1017/S0305000900008667.

96. Oller, D.K., Ramsay, G., Bene, E., Long, H.L., and Griebel, U. (2021). Protophones, the precursors to speech, dominate the human infant vocal landscape. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200255. https://doi.org/10.1098/rstb.2020.0255.

97. Orena, A.J., Byers-Heinlein, K., and Polka, L. (2019). Reliability of the language environment analysis recording system in analyzing French–English bilingual speech. J. Speech Lang. Hear. Res. 62, 2491–2500. https://doi.org/10.1044/2019_JSLHR-L-18-0342.

98. 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. Proc. Natl. Acad. Sci. USA 107, 13354–13359. https://doi.org/10.1073/pnas.1003882107.

99. King, L.S., Camacho, M.C., Montez, D.F., Humphreys, K.L., and Gotlib, I.H. (2020). Naturalistic Language input is associated with resting-state functional connectivity in infancy. J. Neurosci. 41, 424–434. JN-RM-0779-0720. https://doi.org/10.1523/jneurosci.0779-20.

100. Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. Psychol. Sci. 24, 2143–2152. https://doi.org/10.1177/0956797613488145.

101. Cristia, A., Bulgarelli, F., and Bergelson, E. (2020). Accuracy of the language environment analysis system segmentation and metrics: a systematic review. J. Speech Lang. Hear. Res. 63, 1093–1105. https://doi.org/10.1044/2020_JSLHR-19-00017.

102. Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., and Thai-Van, H. (2016). Reliability of the language ENvironment analysis system (LENA™) in European French. Behav. Res. Methods 48, 1109–1124. https://doi.org/10.3758/s13428-015-0634-8.

103. Lynch, M.P., Oller, D.K., Steffens, M.L., and Buder, E.H. (1995). Phrasing in prelinguistic vocalizations. Dev. Psychobiol. 28, 3–25. https://doi.org/10.1002/dev.420280103.

104. Xu, D., Richards, J.A., and Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. J. Speech Lang. Hear. Res. 57, 1638–1650. https://doi.org/10.1044/2014.

105. Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). Analysis of longitudinal data. Oxford Statistical Science Series Oxford Statistical Science Series. https://doi.org/10.1002/(SICI)1097-0258(19960615)15:11<1231::AID-SIM282>3.0.CO;2-Z.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
LENA Automated Analysis of LENA recordings	LENA Foundation: https://www.lena.org/ 361 Centennial Parkway Suite 10 Louisville, CO 80027 303.545.9696	N/A
Deposited data		
The dataset supplying LENA automated analyses of 39,434 all-day recordings, plus a pdf file supplying instructions to generate the GEE analyses	https://data.mendeley.com/datasets/cp7b8vvm38/1	OllerData_Final.xlsx, InstructionsForGEEAnalysis.pdf

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, D. Kimbrough Oller (koller@memphis.edu).

Materials availability

This study did not generate new unique reagents or other materials.

Data and code availability

- Data: Data that yielded the LENA analyses of IVol, CT and AWC are publicly available on Mendeley at https://data.mendeley.com/datasets/cp7b8vvm38/1
- Code: Instructions for running the GEE analyses and SPSS version 28.0 syntax are provided at the same Mendeley site.
- Additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Sample size estimation

There were 5899 infants, 2996 female and 2903 male, sex having been identified by caregivers on enrollment in the programs described below. 39,434 all-day recordings provided the data, with an average of 6.7 recordings per infant.

How subjects/samples were allocated to experimental groups

Girls and boys were allocated to separate groups for analysis, and all-day recordings from each infant were assigned to the age groups of the recordings from 2 to 24 months. The numbers of infants and recordings of each sex at each age are supplied in Table S1 (supplemental information).

General information on participants

The term ''infant(s)'' is used in this article to encompass the age range from 2 to 24 months. 5899 infants participated in recordings from LENA Start and LENA Home, programs of caregiver assistance, both programs constituting caregiver education/early interventions offered by the LENA Foundation. The data were all collected pre-Covid. The programs are described in detail at https://web.archive.org/web/20230308172528/https://www.lena.org/lena-start/ and at https://web.archive.org/web/20230317215424/https://www.lena.org/lena-home/. Briefly, families voluntarily enrolled in one of the programs early in the life of their infants, often within the first months. The goal in both LENA Start and LENA Home was

CellPress
OPEN ACCESS

to help families prepare their infants for school by encouraging, for example, nurturant caregiver-infant vocal interaction and reading together. There was no random assignment of infants to programs; all infants were enrolled for intervention. It is important to emphasize that the present work is not an intervention study; the purpose in combining the data from the two sources is merely to enable analysis of a very large dataset.

All-day recordings were conducted by the families for nominally 6 weeks during each of the programs, and additional educational activities were involved, including regular meetings of participating families with an intervention specialist (for LENA Start only) and/or instructional interactions with families through various media (internet, telephone, and written materials for both LENA Start and LENA Home), and home visits by interventionists (LENA Home only).

It is important to emphasize that since these LENA recordings were the product of intervention efforts, conducted all over the USA in numerous programs of family assistance for infant and child development, there was no practical possibility to systematically incorporate such desirable research characteristics as detailed questionnaires about home life, number of siblings, socio-economic status, ancestry, race or ethnicity, nor was there any possible control group or opportunity for random assignment of infants to groups. Of course many prior research projects conducted with the LENA recorder and analysis system and by collaborators of the LENA Foundation have indeed included design features allowing greater access to demographic data and flexibility of comparisons (see https://web.archive.org/web/20230317215403/https://www.lena.org/research/ and https://web.archive.org/web/20230317215425/https://www.lena.org/wp-content/uploads/pdf/LENAUserPresentations.pdf for a list of scores of relevant journal publications and formal presentations).

All participant caregivers signed a consent form that allowed data from the automated analysis to be preserved and analyzed for publication, but unlike many prior LENA studies, did not allow preservation of the recordings themselves, all of which were erased after automated analysis. Consequently, the analysis that can be presented here is limited to the automated outcomes of the standard LENA algorithms—it offers an opportunistic perspective on the CB-2020 result, but it is restricted in interpretive options.

The 2996 female and 2903 male infants in the study were distributed 54% Start/46% Home. We have only one demographic variable that can be very roughly estimated. The socio-economic status (SES) of the families can be estimated roughly based on proxy data available on the Area Deprivation Index (ADI) for 32% of the participants, whose residences were scattered across the US, from coast to coast (see https://web.archive.org/web/20230317215404/https://www.lena.org/where-are-lena-programs/). Using the small amount of ADI data available, the average for infants in the study can be estimated at the 38th percentile, suggesting relatively low SES, consistent with the fact that the goal of the intervention programs was to help accelerate language development in infants who might otherwise begin school at a disadvantage.

METHOD DETAILS

Numbers and lengths of recordings

39,434 all-day recordings provided the data, with an average of 6.7 recordings per infant. The data were semi-longitudinal with some infants beginning their participation by as early as the first months of life and ending usually within six weeks but sometimes up to several months after starting. The data to be presented here involved recordings supplied by infants who began participation at any age through 24 months. The entire corpus to be presented here included data from 2 through 24 months as portrayed in Table S1 (supplemental information). Optimally, weekly recordings were supplied, but many infants supplied fewer and even in some cases, just one all-day recording (see distributions in Figure S1). An average of over 1700 recordings were supplied for analysis at each of the ages, with an average of more than 800 infants supplying recordings at each age, roughly evenly distributed between boys and girls.

The original recordings were sometimes as long as 16 hours. All recordings were trimmed to not more than 12 hours and not less than 10 hours, based on elimination of periods beyond 12 hours and in a few cases, beyond 10 hours—mean recording length was 11.97 hours. By targeting 12-hour durations, it was possible to achieve maximal comparability across recordings in the present work and from other studies, where 12-hour durations are often standard. Furthermore, all three parameters were computed to yield a value

CellPress
OPEN ACCESS

iScience
Article

per minute (IVol/min, CT/min, and AWC/min) within the trimmed recording period to enhance comparability across samples of different lengths.

Recording device and automated analyses

The LENA system includes a small battery-powered audio recorder worn on the infant chest in special clothing. For more information on the device including photographs of the recorder and some of the available clothing go to https://shop.lena.org/. The LENA automated analysis of such recordings yields measures of IVol (called "Child Vocalizations"), CT, and AWC directly for all end users.

In brief, the automated system is based on 8 Gaussian mixture models trained on a substantial corpus of human-coded real speech from LENA recordings in homes of children from 2 to 48 months of age. The primary categories assigned by the algorithm include voices of: 1) the infant or child wearing the recorder (CH), 2) any other infant or child (CX), 3) any male adult (MA), 4) any female adult (FA), 5) television or other electronic voices (TV), 6) overlapping sounds (OL), i.e., simultaneously occurring combinations of categories 1 through 5, 7) noise not associated with voice (NO), and 8) silence (SIL). The algorithm assigns one of the categories for each segment of the recording, always comparing the likelihood of categories 1–7 with silence and selecting the most likely option if any of them differs from silence above a predetermined threshold of probability. Extensive description of the system can be found in the LENA Technical Reports, https://web.archive.org/web/20230317215404/https://www.lena.org/technology/#tech-reports, where data on the hardware and software as well as validation data for the automated coding are presented. Additional validation data and reviews can be found in a variety of articles.[32,33,44,97–102] The measure of infant volubility (IVol) is based on protophones and speech occurring in category 1 above (CH). CH also includes segments of crying and vegetative sounds (such as coughing or sneezing), but these are not included in the counts we used. The CH measure is designed to provide a basis to derive counts of protophones and speech "utterances", which are defined as vocal breath groups (utterances) of the infant/child.[103] The measure of conversational turns (CT) is dependent on the concept of "conversational blocks" as implemented in the LENA software, where a sequence of utterances produced by any speaker (except for TV voices) constitutes a block if there is no interval of silence greater than 5 seconds within the sequence. A CT is counted whenever a category 1 (infant) protophone or speech utterance immediately precedes or follows a category 3 (male adult) or category 4 (female adult) utterance within a conversational block. Adult word count (AWC) is estimated by the LENA algorithm based on a standard phoneme counting model.

The fact that there is a wide range of levels of agreement between human coding and the LENA measures reported in the various studies cited above does not provide a reason to doubt the statistically significant results reported in the present paper. The agreement studies routinely report *statistically significant positive associations* between human coding and the LENA automated outcomes for all three variables. This fact implies that there is a statistically significant signal to assess in all three cases (i.e., statistically significant and thus statistically reliable LENA measures of IVol, CT and AWC). The danger for research where agreement between the gold standard (human coding) and the measure of focus (LENA outcome) is low in some absolute sense (for example, *r* = .3) is that the low agreement contributes to low power and possible Type II error, i.e., failure to find a real difference. We have expected to be relatively protected against Type II error in the present work by having a sample size of thousands of infants and tens of thousands of recordings. But statistically significant differences (boys more voluble than girls at some ages, for example) even in cases where agreement is low remain valid just as any real signal detection in noise is valid.

No original recordings available

It is important to reiterate that we had no access to the original recordings and thus could not and cannot perform human coding on the sample. Human coding remains the gold standard for counting vocalizations in recorded samples, but extensive comparisons of the results of LENA automated analysis with human-coded samples suggests that for all three measures evaluated here, there is good reason to take the obtained automated results seriously (for empirical verification, see prior publications[44,49,98,104]).

QUANTIFICATION AND STATISTICAL ANALYSIS

Generalized Estimating Equations (GEE)[46,105] provide a conservative, non-parametric method for analyzing complex longitudinal data. Fixed and random effects can be accounted for, but GEE is preferable to

traditional mixed models whenever there are correlations among data from participants across conditions (especially in this case across ages), and when the number of observations varies for participants within or across conditions (for example, the number of recordings per infant varied from 1 to more than 20 in the present work). The GEE approach has the advantages of requiring no normality assumption and being relatively invulnerable to overestimating differences across groups when participants in the groups are distributed in substantially different ways.

To find the original data for this article go to Mendeley Data with the link provided above under Deposited data and find OllerEtAL_LENA. The data file is OllerData_Final.xlsx. There is a pdf instruction file (InstructionsForGEEAnalyses) at the same site indicating how to run the GEE analyses in SPSS version 28.0.