Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study

Alejandrina Cristia, and Emmanuel Dupoux
LSCP, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University

Michael Gurven
University of California at Santa Barbara

Jonathan Stieglitz
Université Toulouse 1 Capitole

This article provides an estimation of how frequently, and from whom, children aged 0–11 years (*N*s between 9 and 24) receive one-on-one verbal input among Tsimane forager-horticulturalists of lowland Bolivia. Analyses of systematic daytime behavioral observations reveal < 1 min per daylight hour is spent talking to children younger than 4 years of age, which is 4 times less than estimates for others present at the same time and place. Adults provide a majority of the input at 0–3 years of age but not afterward. When integrated with previous work, these results reveal large cross-cultural variation in the linguistic experiences provided to young children. Consideration of more diverse human populations is necessary to build generalizable theories of language acquisition.

Language is ubiquitous in human cultures, yet talkativeness is thought to vary across cultures in ways that affect the prevalence and format of child-directed speech (Richman, Miller, & LeVine, 1992). Based on the assumption that a child's early experiences with speech may impact their language acquisition, the present work aims to document the prevalence of speech addressed to children in a forager-farmer population. In this Introduction, we first lay out the conceptual motivation of this line of research and review extant literature on the topic before introducing the methods and questions explored in the present study.

Causal Pathways Between Child-Directed Speech Quantities and Acquisition Outcomes: Theory and Data

From a conceptual viewpoint, amount of adult speech directed to infants has been proposed to impact their language development in two main ways. First, if more speech is addressed to the infant, then (all else being equal) he or she has more opportunities to hear the forms of words in a meaningful social context, which may facilitate learning of word-to-meaning pairings (Hoff & Naigles, 2002). Adults are presumably more skilled than children in adapting their speech to their interlocutors' needs (Davies & Katsos, 2010; Street & Cappella, 1989), and so adult interlocutors may provide more useful contextualized verbal input than child interlocutors. Second, getting comparatively less directed input during early childhood, and thus having fewer learning opportunities, may result not only in lower levels of explicit knowledge (e.g., smaller vocabularies) but also less efficient speech processing (Weisleder & Fernald, 2013), potentially as a side effect of developing less robust lexical and phonological categories. This effect could magnify the previous one, as low-input infants would be slower to process information from the little speech they do hear.

There is considerable empirical support for the broader theoretical view that child-directed verbal input from adults shapes children's language development, particularly when the latter is measured through receptive or productive vocabulary size as well as speed of word recognition. In a benchmark study, Hart and Risley (1995) argued that differences in verbal skills (and ultimately academic abilities) between children of varying socioeconomic status (SES) could be traced back to the verbal input directed to these children between birth and 3 years of age. In quantitative terms, those authors estimated that children of professional parents heard *three times as many words* than same-aged peers growing up in households on welfare. Other studies have documented the predictive value of quantity of speech directly addressed to children aged 1–3 years with respect to concurrent or subsequent linguistic development even when controlling for SES (e.g., Hoff & Naigles, 2002; Rowe, 2008; Weisleder & Fernald, 2013; see also Hirsh-Pasek et al., 2015; Rowe, 2012; Rowe & Goldin-Meadow, 2009; Cartmill et al., 2013 for additional considerations; and Hodson, 2014; Baker-Henningham & López Boo, 2010 for discussion of intervention studies). Evidence from diverse societies suggests that child-directed input is more effective in promoting verbal engagement and lexical development than overheard speech (Shneidman, 2010; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2013), particularly when uttered in one-on-one conversations (Ramírez-Esparza, García-Sierra, & Kuhl, 2014) with adults rather than children (Shneidman et al., 2013).

Much of the evidence on early language development comes from the study of a small number of populations, which could be described as "Western, educated, industrialized, rich, and democratic" (WEIRD; Henrich, Heine, & Norenzayan, 2010). Such populations may not be representative of human psychology at large, both because they are a biased sample of today's population and because this particular combination of characteristics is rare in evolutionary terms. For most of their evolutionary history, humans have foraged for their own food, which has a host of consequences regarding group size and organization, and ultimately infants' language experiences. Compared to industrialized populations, preindustrial populations have larger families and tend to live in smaller, kin-based clusters, often sharing living spaces, whereby infants regularly come into contact with extended kin. These differences may both increase the number and diversity of potential conversational partners, and decrease the number of one-on-one conversations in which the infant is engaged. Much current work on verbal input to infants focuses on a single adult who is the child's primary caregiver, often the mother (e.g., Hoff & Naigles, 2002; Rowe, 2008). Might these situations also be prevalent in preindustrial societies? On one hand, adults in preindustrial societies need to spend more of their time on food production or general maintenance tasks than children, who are weaker and less skilled, and therefore spend less time in productive activities. Fewer opportunity costs mean that older children have more time to spend interacting with, and potentially speaking to, their younger peers (Stieglitz, Gurven, Kaplan, & Hooper, 2013; Weisner & Gallimore, 1977). On the other hand, as mothers in preindustrial societies tend to practice on-demand breastfeeding, infants may spend the majority of their time around the mother and thus in a position to receive speech mostly from her until they are weaned, which occurs relatively late (on average at 2.5 years in a variety of preindustrial societies; Sellen & Smay, 2001).

Previous Work on Child-Directed Speech in Preindustrial Societies and Comparison Groups

Two waves of empirical studies bear on the quantities and sources of infant-directed speech. The first consisted mainly of anecdotal observations, which by and large supported the idea that, in preindustrial societies, infants are rarely observed in deliberate dyadic verbal interactions, particularly with adults (reviewed in Lieven, 1994). To date, we have found six empirical studies that report precise, systematic estimates of child-directed verbal input quantity and/or sources based on recordings and/or observations of spontaneous interactions involving young children, which we summarize next.

Three of the studies used systematic observations: Infants aged 8–16 months and their caretakers were observed for 15–20 min at a time (depending on the study), and the presence or absence of vocalizations, together with a few other parameters (e.g., distance between caretaker and child), were coded every 5 s (Klein, Lasky, Yarbrough, Habicht, & Sellers, 1977; Konner, 1977; Tulkin & Kagan, 1972). Results for 10-month-olds in preindustrial populations were as follows: The proportion of 5-s segments containing caretaker vocalizations was 4% among 20 dyads observed in a Guatemalan village (Klein et al., 1977) and 10% for 9 dyads observed in a !Kung hunter-gatherer camp (Konner, 1977). The latter percentage is, in

fact, higher than that observed among 26 working-class dyads in Boston, for whom 7% of segments contained vocalizations, but slightly lower than that found for 30 middle-class Bostonian dyads, namely 15% (Tulkin & Kagan, 1972; see also Tulkin, 1977).

More recently, Vogt, Mastin, and Schots (2015) videotaped 14 children in rural Mozambique, urban Mozambique, and the Netherlands, at 13 and 17 months, whereas Shneidman and Goldin-Meadow (2012) did so for 6–9 Mayan children and 9 children from Chicago. Results from both studies are summarized in Table 1. Focusing first on the ratios of quantity of directed speech experienced by children growing up in preindustrial, as compared to industrial, settings, it is clear that the former hear a great deal less directed speech, particularly at early ages. Shneidman's data suggest that, to a certain extent, this may be explained by overall differences of talkativeness across the cultures: Notice that when quantity of directed speech is encoded in terms of overall proportion of the input that is directed to the child, then cross-cultural differences in this study are greatly reduced by the time children are about 30–33 months of age. We return to similarities and differences across studies in the Discussion.

Systematic evidence on who talks to children is less abundant. To take a specific example, Harkness (1977) observed and audiotaped 20 children between 2 and 3.5 years of age in a rural settlement for Kenyan Kipsigis families. These target children were observed with fellow children about 75% of the time, with their mother 50%, and other adults about 25%. (Notice that these categories are not mutually exclusive: A child may be with the mother, other adults, and other children at the same time.) However, this may not translate into higher volumes of input spoken by fellow children, given

that a higher proportion of time spent with adults involved language than time spent with fellow children. Unfortunately, Harkness (1977) does not quantify speech received from children versus adults. Only one study quantifies who speaks to children (see Mastin, 2013 for attention and interaction). Shneidman (2010) documents dramatically different patterns in the Mayan and American recordings. Among Mayan 13-month-olds, 60% of all sentences produced around the child (collapsing across directed and overheard) came from other children (defined as individuals under 11 years of age), 31% from the mother, and 9% from other adults. In the case of American 14-month-olds, only 8% of all sentences (collapsing across directed and overheard) were uttered by fellow children, 79% by the mother and 13% by other adults. In a separate longitudinal study, Shneidman (2010) found that the percentage of all sentences (collapsing across directed and overheard) from a child source increases with the target child's age, reaching 90% by 3 years among Mayans, whereas it remains stable at about 10% for Americans. A similar picture ensues when one focuses on child-directed sentences: At 13–14 months, the mother contributes 24% of the directed input for the Mayans versus 87% for the Americans, other adults about 11% in both, with fellow children providing the remainder (roughly 65% for the Mayans and 1% for the Americans). The proportion of directed speech coming from the mother is stable at around 19%–33% for Mayan children observed at 18–35 months of age (L. A. Shneidman, personal communication, 2017-08-24). As for the other studies discussed in the frequency/quantity section above, none of them explicitly breaks down vocalization frequency/quantity as a function of who speaks.

Table 1

Summary of Results Reported in Vogt et al. (2015) and Shneidman and Goldin-Meadow (2012)

Pre-ind.	Age	OH	Dir.	%	Ind.	Age	OH	Dir.	%	Ind/Pre
Rural Mozambique	13		43		Dutch	13		436		10
Rural Mozambique	17		107		Dutch	17		671		6
Urban Mozambique	13		207		Dutch	13		436		2
Urban Mozambique	17		243		Dutch	17		671		3
Mayan	13	220	55	20	Chicago	14	341	605	64	11
Mayan	24	262	228	47	Chicago	23	475	652	58	3
Mayan	33	142	209	60	Chicago	30	631	970	61	5

Note. Each study reported number of utterances directed to children (Dir.) at a range of ages (in months), from both a preindustrial (Pre-ind.) population and a comparison industrial population (Ind.). Shneidman also reported number of utterances overheard by the child (OH), allowing the calculation of the percentage of input that is directed to the child out of the total represented by summing overheard and directed. Ind/Pre indicates the ratio of number of utterances directed to children in industrial settings divided by number of utterances directed to children in preindustrial settings.

Present Work

This study addresses the question of whether infants in a preindustrial society, Tsimane forager-horticulturalists of lowland Bolivia, receive little directed input from adults using time allocation, an observational technique for systematically monitoring behavior (Gross, 1984; Johnson, 1975; Johnson & Behrens, 1989; Mulder et al., 1985). Observers were not specifically targeting speech but coding numerous behaviors, rendering unlikely that theoretical biases regarding language development affect coding. Furthermore, because participants were observed for all behaviors and not just verbal output, it is unlikely that they changed their own behavior on the basis of, for instance, their expectations of what the researchers wanted to observe in this domain (see Supporting Information, https://osf.io/jz2u5/, Section 1). In contrast, most previous studies used ostensive recording equipment, and/or the focus of the study was language development or verbal behavior. It is possible that these factors affected the data (for instance, see Shneidman & Goldin-Meadow, 2012). Given these considerations, we suggest that the present data provide a useful complement to more targeted observations and video recordings such that if our results are similar to those found with other methods and other populations, researcher bias and observer effects are less likely to be a concern for the body of literature as a whole.

Additionally, previous conclusions about populations were based on samples that were not necessarily representative of the population, often observing individuals only once, and a narrow range of ages (with a focus on 8 months to 2 years of age). In the present study, residential clusters were sampled from six representative villages, in which about 70,000 observations of residents and visitors aged 0–85 years were recorded during 2- or 3-hr blocks, at times ranging between 7 a.m. and 7 p.m. This systematic, large-scale, and representative coverage in a preindustrial population is a welcome addition to extant work describing children's verbal input.

Method

This project has been documented using the Open Science Framework. The link to the elements that could be rendered public, namely scripts, derived data, and reports, is available from https://osf.io/5bjs6/. The direct link to the Supporting Information is https://osf.io/jz2u5/.

Study Population

Tsimane inhabit over 90 villages ranging in size between 50 and 550 individuals. They cultivate plantains, rice, corn, sweet manioc, and other crops in small swiddens, and regularly fish and hunt. At the time of data collection, these foods comprised more than 90% of the diet, with the remainder purchased from market stores or obtained from trade with itinerant merchants. Villages are composed of extended family clusters (each containing about three or four households), where the majority of food and labor sharing occurs. Communication typically occurs in the native Tsimane language, one of three dialects of the Mosetenan language family (Campbell, 2012); Spanish may be spoken to non-Tsimane Bolivians (e.g., merchants). Women have their first child by 19 years of age, on average, with an interbirth interval averaging 30 months (Stieglitz et al., 2015), and a total fertility rate of about nine births (Kaplan, Hooper, Stieglitz, & Gurven, 2015). Infants are kept close to their mothers, and regularly carried in a sling so that mothers can perform subsistence activities; toddlers are often cared for by older siblings or other kin. A recent study (using the same data set analyzed here) showed that mothers provide 80% of the direct child care in the first 6 months of life, and 70% in the first 6 years (Winking, Gurven, Kaplan, & Stieglitz, 2009).

Data Set

Data were collected as part of the University of New Mexico–University of California Santa Barbara Tsimane Health and Life History Project (http://www.unm.edu/~tsimane/). Demographic and genealogic data were derived through a combination of methods, for example, by interviewing family members and, if possible, cross-checking this information with official logs (see Gurven, Kaplan, & Supa, 2007 for more detailed explanations). Behavioral observations following Johnson (1975) were conducted in 2002–2003 in four communities and in 2005 in two more communities. In each community, multiple residential clusters were defined, and a cluster of households (usually 3–4) was sampled (without replacement) to be observed for a period of 3 (2002–2003) or 2 (2005) hours at a time.

During the visit, a single observer was in a position where he or she could observe without interfering with the activities of cluster residents. Every 30 min, the observer coded up to two concurrent activities (coding also location and interactant, if relevant) of each resident and that of visitors

present at the time. Which two activities were coded depended on the observed person's focus of attention. For example, imagine an 11-year-old girl who (a) climbs a tree, (b) talks to her friend, and (c) passively boils plantain over the fire. Regarding (c), the pot of plantains sits on the fire, without the girl actively watching, stirring, adding/removing plantain, or adjusting the fire. In such a case, the first activity would be "climb tree," and the second "talks to friend," while omitting the third activity "passively cooking plantain." Observers were advanced anthropology students (working toward their PhD or an honors thesis). They resided in Tsimane villages, visiting families regularly for some time (weeks or months) prior to collecting any time allocation data. Observers also studied Tsimane language (e.g., written orthographies, common phrases) before even going to villages and regularly interacted with Tsimane research assistants, who could help clarify language-related questions, before and during data collection. Given the granularity of analyses presented below, all the observer needed to do was detect the presence versus absence of speech—something that any speaker can do even in a non-native spoken language. "Speaking" was coded as an activity without a specific request that at least or at most a certain quantity of speech was uttered. Observers were not given specific instructions regarding how to decide whether speaking involved a single interactant or multiple ones, with the exception of one category of verbal behavior that was reserved for cases in which the conversation involved three or more participants from at least two households. In all other cases, it was left to the observer to decide who the interlocutor(s) were, which they could do based on verbal cues (use of the interlocutor's name, content of the conversation) and nonverbal cues (including gaze, as Tsimane usually look at each other when speaking). Physical proximity was not required, and as Tsimane houses often lack walls, interactants could also be in different locations (e.g., the yard and the kitchen). There was only one slot for a possible interactant; thus, all speech directed to a group would have been coded without a single interactant, and—as explained below—will be counted as undirected speech.

We call a *scan* a single unique observation: the conjunction of one individual observed in a given cluster, on a given date, at a given time. We call a *slice* the group of scans that have been gathered in a given cluster, on a given date, and at a given time, and which, therefore, pertain to any number of residents. Thus, one visit to a cluster contains four or six slices (based on whether clusters were visited for 2 or 3 hr), with a scan conducted every half hour, and $4 \times N$ or $6 \times N$ scans, with N being the number of people whose activity is being noted in that visit. In total, the data set contains about 70,000 scans. This contains data for all residents of a cluster, even if absent, as well as visitors. Given our interest on a behavior that is transitory and may not be easily remembered or reported as pinpointed at one specific time, we rely on the 43,903 observations where observers directly observed participants. All 43,903 direct observations, each containing information of up to two activities (each potentially involving an interactant) for each person in each slice, were used in subsequent analyses. These observations covered all six communities, containing between 4 and 21 clusters each ($M = 9.5$). There were a total of 3,854 slices, containing 1–49 individuals each (median = 10).

Data Processing and Analyses

All analyses were carried out in R (R Core Team, 2015) and rendered in this manuscript using knitr (Xie, 2014) and xtable (Dahl, 2009). The goal of these analyses was to calculate estimates of time that people spend talking with those in a given age group. Given the way that these estimates were calculated, they should be viewed as "observation-weighted" aggregate measures over an age group not averages across individuals.

The first step of processing involves determining who is counted as part of the focal group. These are people who have at least 50 scans in the age range being considered, which means that they, and those around them, were observed for a minimum of eight separate visits (as 50 scans are drawn from 25 observation hours). As we were computing observation-weighted means, we could have relaxed this criterion. However, having representative data from individuals allowed us to further calculate 95% confidence intervals across individuals, using a method explained below. The age ranges considered were a function of our scientific interests, which led us to define fine-grained distinctions within infancy: 1 year wide from birth to 4 years of age, 2 years from 4 to 8, and 3 years from 8 to 11. As shown in Table 2, our conclusions are based on considerable numbers of scans: Focal age groups included 9–24 individuals, and these individuals were surrounded by many others, such that there are 366–638 unique individuals who could have potentially spoken to those in the focal group at a given age range. The "Avg in location" column codes, for a given scan

Table 2
General Characteristics of the Samples Defined by Each Age Group

Age range (years)	Focal (%f)	F-scans	Slices	Others	O-scans	Avg in location
0–1	24 (33)	1,546	1,185	636	16,649	6.62
1–2	9 (56)	649	565	366	8,299	7.71
2–3	13 (62)	882	769	425	12,350	7.63
3–4	25 (40)	1,750	1,329	617	15,857	5.88
4–6	26 (58)	2,104	1,567	606	20,285	6.05
6–8	23 (13)	1,587	1,192	512	16,189	6.11
8–11	32 (59)	2,105	1,525	688	20,660	6.11

Note. Age group (range, in years); number of people who could be included in the focal group (percentage of women, %f); number of scans with people in the focal group as agents (F-scans); number of slices in which those focal people were observed (Slices); number of other people whose data were collected in those same slices (Others); number of scans with these other people as agents (O-scans); Avg in location indicates in each slice involving a focal person, how many other people are present, on average, in the same precise location as the focal person.

involving a person in the focal group, the average number of scans involving other people that were made at the same precise time and location (i.e., the kitchen or the yard). Inspection of this column indicates that the number of people in close physical proximity to those in the focal group is stable across age groups. In supplementary analyses, we ensured that all seasons (wet, which is roughly November–April; dry, roughly May–October) and daylight hours were similarly represented across age groups, and that our conclusions below held for children growing up in the most acculturated village versus other less acculturated villages, as well as across genders (for more information, see Supporting Information, https://osf.io/jz2u5/, Sections 2–4).

The rest of the procedure is illustrated in the flowchart in Figure 1 for the age range between birth and 1 year of age as an example. Once the people in a focal group are identified, we determine in which slices these focal people have been observed using cluster number, date, and time, and we then use this information to identify all scans included in those slices, both those where people in the focal group are agents and those where others are agents. As noted above, each scan contains up to two activities, each with a potential interactant. We then identified all scans containing speech as either (or both) of the activities (see also Supporting Information, Section 5, for further methodological information). Depending on the identity of the interactant, we classified scans in which speech was observed as follows:
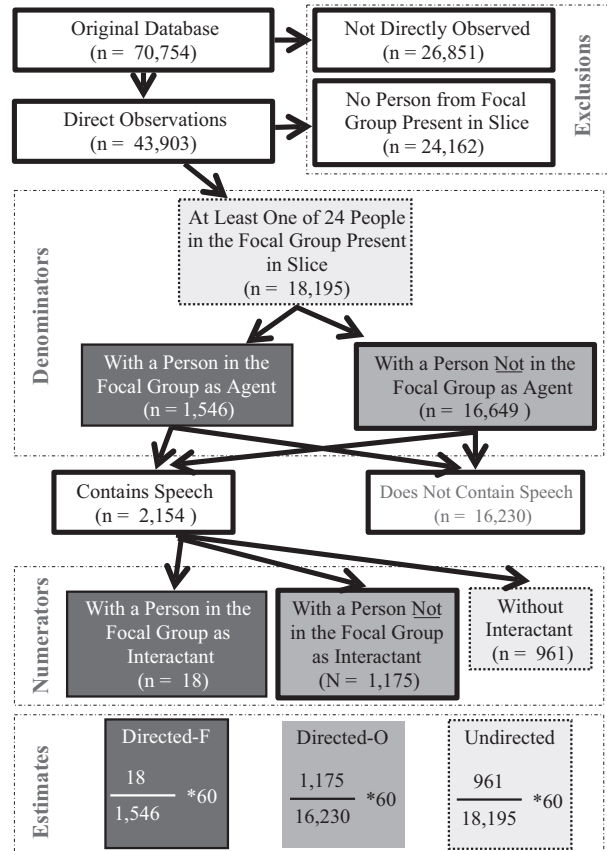

Figure 1. Flowchart representing scan inclusion and exclusion when infants between birth and 1 year of age are defined as the focal group, in order to estimate speech that was not coded as involving an interactant (Undirected), speech that was directed to infants in the focal group (Directed-F), and speech that was directed to other people who were observed in the same slices as those infants (Directed-O). The color coding helps connect numerators and denominators contributing to the three estimates.

1. If the personal identification code for the interactant is one of the focal people, then this scan counts toward "one-on-one speech directed at people in the focal age group" (Directed-F).
2. If the personal identification code for the interactant is *not* one of the focal people, then this scan counts toward "one-on-one speech directed at people not in the focal age group" (Directed-O).
3. If the personal identification code for the interactant is blank, then this scan counts toward "undirected speech" (Undirected).

In order to convert these counts (e.g., number of scans containing Directed-F speech) into time estimates, it is necessary to calculate proportions where these counts are the numerator and the

denominator is the maximum total possible number of scans. Let the number of scans where people in the focal group are agents be N_FOC; the number of scans where people not in the focal group are agents be N_NONFOC; these two add up to N_TOTAL. Given that a person can only have a one-on-one directed conversation with one other person at a given time, the maximum total possible for Directed-F is N_FOC, and for Directed-O is N_NONFOC. In contrast, the denominator for Undirected is N_TOTAL because in principle all scans (N_FOC + N_NONFOC) could contain undirected speech. Finally, these ratios are multiplied by 60 to estimate the number of minutes per daylight hour each type represents. We note here that this is an estimation in terms of time spent in the relevant activity (e.g., amount of time speaking to infants) and not in terms of quantity of speech (utterances, words) produced.

To provide estimates of the variance in our estimations, we used bootstrap resampling to derive 95% confidence intervals within each age group. This was done by sampling (with replacement) from the children in each focal age group 10,000 times and extracting the 2.5 and 97.5 percentiles in the resulting distribution of the three dependent variables (see Supporting Information, Section 5C for a more detailed explanation).

Results

Our estimations from the time allocation data revealed that < 1 min per daylight hour is spent talking to children below 4 years of age (Figure 2). This is much lower than the amount of time spent talking to other people present in the exact same

slices where young children were observed, which is about 4 min per daylight hour. Estimates of the amount of time spent talking to children binned in 1-year intervals from birth to 4 years of age are very similar and within 10%–12% of each other. Higher estimates are observed when analyses focus on older children as interactants: roughly 2 min per hour among children aged 4–8 years, and about 4 min per hour among children aged 8%–11 years.

In view of the relatively low prevalence of speech directed to young children, analyses pertaining to *sources* of directed speech are based on broader age ranges. Indeed, even though there were 24 children who had at least 50 scans as agents when they were between 0 and 1 year of age, they were coded as receiving directed speech in only 18 instances out of a total possible maximum of 1,546 scans (Figure 1). To have larger samples of scans from which to draw stabler conclusions, we broadened the early age ranges to 3 years. Figure 3 shows that the single greatest producer of speech directed to infants is the mother. Fellow children, and most saliently brothers and sisters under 12 years old, provide about 38% of the input addressed to children younger than 3 years of age. As a result, the majority of child-directed speech comes from adults in these early years. However, fellow children provide growing proportions of input in middle childhood. Other adults contribute directed input relatively seldom at all ages studied here.

Discussion

Using a time allocation data set in a preindustrial society, we found that Tsimane spend little time
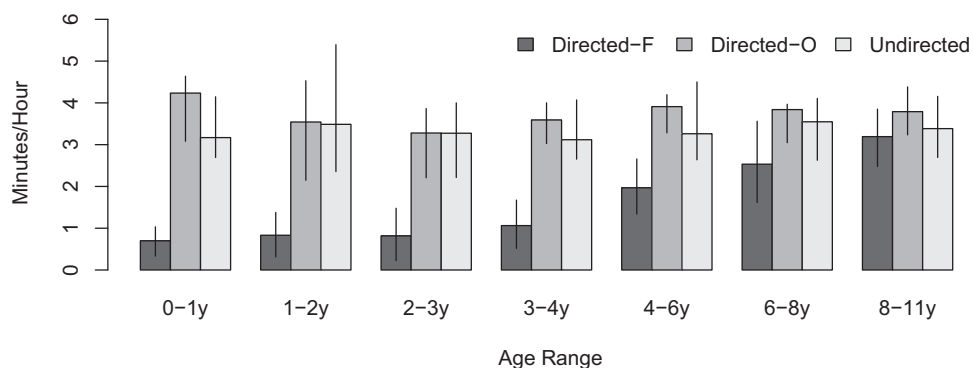

Figure 2. Frequency of speech directed to children in each focal group (Directed-F), directed to other people in the same slices (Directed-O), or Undirected, as a function of focal age range. Bars indicate 95% confidence intervals estimated using bootstrap resampling over individuals.
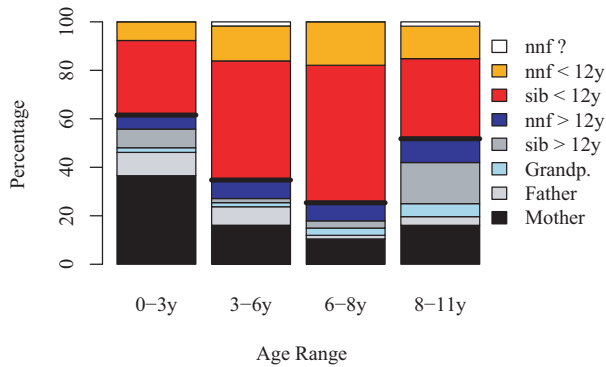
Figure 3. Sources of directed verbal input as a function of age of the focal group. Sources are ordered as a function of age (adults below the thick line, children above it) and relationship. "Grandp" stands for grandparents, "sib" for siblings, "nnf" for not in the nuclear family. "?" indicates that the speaker's age was not known (which sometimes occurred for visitors who were not part of the study).

speaking to infants and young children. We state that these frequency estimates are *low* based mainly on the comparison between this estimation and similarly calculated estimates of time spent talking to others present in the exact same slices where the young children were observed, which, given group composition, represents a sort of average of amount of speech directed to people of various ages. It is also clear that the same analyses focused on slightly older children reveal a rather different picture: The prevalence of one-on-one speech with children aged 8–11 years as interactants is more similar to the other directed amounts, with largely overlapping confidence intervals across the two. In other words, it is not the case that speech, including one-on-one speech, is particularly rare among Tsimane families, as substantially higher estimates were found for speech directed to others in the same visits where infants were observed, as well as in analyses focused on older children. Rather it appears that infants and young children *specifically* are more seldom engaged in one-on-one conversations than older individuals. The sources of child-directed speech, meaning the people who actually talk to the young children, were very different as a function of children's age: Mothers were the main contributors between birth and 3 years of age, and overall a statistical majority of speech to infants comes from adults. In contrast, the majority of speech observations involving older children as interactants had fellow children as speakers.

In the remainder of this section, we integrate our results with previous research, draw implications for current theories of language acquisition, discuss

a number of limitations to these results, and conclude by identifying some open questions.

Integrating Present Work With Previous Research

Given the variability in methods found across studies, this section aims to juxtapose our results to those of previous work as much as possible, as this article would be incomplete if we did not attempt this integration. We start from the conclusion that Tsimane infants and young children receive relatively little one-on-one directed spoken input from adults, with "relatively" most accurately being interpreted relative to others in the same culture. As noted in the Introduction, it is rare that studies report quantities of speech addressed to others in the same culture, so we cannot contrast how much less of an addressee Tsimane infants are compared to older partners, with the same age-based comparison in other studies. However, we can attempt a comparison with other studies that also employed observations (Klein et al., 1977; Konner, 1977; Tulkin & Kagan, 1972), all of which focused in infancy. To perform this integration, we multiplied the percentage of 5-s segments in which vocalizations were observed (the measure reported by this previous work) by 60 to provide an estimate of number of minutes per hour, more similar to our own results. We further selected from our analyses those focusing on infants between birth and 1 year of age (similar results ensue for 0–2 years). As apparent in Table 3, the estimate for Tsimane is lower than all other estimates. There could be three methodological reasons for this divergence. First, all other studies on this table counted presence of vocalization if some vocalization was present at any point within the 5-s segment, whereas our

Table 3

Quantity of Caretaker Vocalization Directed to Infants Reported in Previous Work Using Systematic Observations (See Main Text for Details), Converted From Percentage of Observations Into Minutes Per Hour to Facilitate Comparison With Current Results

Citation	Population	Age (months)	Min/hr
This study	Tsimane	0–12	0.7
Klein et al. (1977)	Guatemala	8	2
Klein et al. (1977)	Guatemala	12	3
Tulkin and Kagan (1972)	Boston Working	10	6
Klein et al. (1977)	Guatemala	16	6
Konner (1977)	!Kung	10	6
Tulkin and Kagan (1972)	Boston Middle	10	10

observers were instructed to code behavior during the instantaneous scan. Second, all previous work explicitly focused on verbal behavior (in the context of infant attachment theory, Bretherton, 1985) and thus coded only a handful of behaviors, whereas our observers were asked to report on all behaviors, with a maximum of two activities at a given point in time. Third, we focus on one-on-one speech, whereas it is unclear from previous descriptions whether undirected (e.g., speaking to the child as part of a group) were also included in the same counts. It should be mentioned that there is another source of divergence across our study and previous ones, although going in the opposite direction: We include vocalizations from all people who addressed the child and not only mothers (Tulkin & Kagan, 1972) or caretakers (possibly what Konner and Klein did). It would be extremely interesting for future work to use a single method to compare across cultures, so as to be better able to assess the extent of cultural differences in amount of speech directed to infants.

We now turn to another aspect of our results, namely comparisons across ages. Inspection of multiple age groups in the Tsimane data revealed that estimates of directed, one-on-one speech quantity were remarkably similar for interactants between birth and 4 years of age, and that the amount of time speaking to children was similar to that spent talking to others only when analyzing much older children, aged between 8 and 11 years, as interactants. These results seem to suggest differences when compared to previous work in preindustrial societies mentioned in the Introduction, which has documented increases in input quantity when comparing 17- or 30-month-old infants against 13- or 14-month-olds (Shneidman, 2010; Vogt et al., 2015).

We first dispel two notions that could appear as alternative explanations but do not account for the observed empirical patterns. First, the broad age ranges used in the present work do not explain why we fail to find an age-related increase in the first 2 years of life. Our reading of the previous literature is that quantity increases are not thought to be transitory, and thus it appears unlikely that we fail to find increases in the second year of life because we are averaging a peak at months 17 and 24 with troughs at the remaining months. Second, this divergence cannot be due to methodological differences across papers. Indeed, researchers applied the same methods to attempt to quantify speech prevalence in all the age groups studied, thus allowing us to contrast across age-based

comparisons of like data. Naturally, any such broad measurement has limitations. Counting amount of time in the activity or number of sentences is insufficient to detect other potential changes in children's input with development, such as in number of word tokens, sentences' syntactic complexity, or overall lexical diversity. Thus, our focus here is on broad estimates of frequency of speech.

In fact, inspection of previous literature had already revealed a diversity of developmental trajectories with such broad quantity estimations (see Table 1): In some societies, young infants are spoken to a great deal and either there is little change with age (American) or they receive even more speech when they are older (Dutch); in other societies, infants receive (relatively) little directed input when very young, and later they may receive a great deal more (Mayan, rural Mozambique, Guatemala) or only somewhat more (urban Mozambique).

Factors Probably Accounting for Variation Within and Across Cultures

The evidence on who talks to infants is extremely scarce at present, with the two extant studies (Harkness, 1977; Shneidman, 2010) showing that most speech comes from adults rather than children, and the mother in particular, and thus we do not discuss variation along this dimension any further. An integration of extant work on (pre)industrial societies suggests that there are marked cultural differences in quantity of infant-directed speech, as well as different developmental curves in terms of quantity of directed speech as a function of the child's age. Further work employing homogeneous, cross-culturally appropriate methods is needed to more accurately measure the extent of this variation and to make strict comparisons possible. Therefore, we do not attempt direct comparisons but rather explore the promise of a few potential factors underlying variation in caregivers' verbal behavior.

A good candidate factor is SES, that is, familial or community differences in a set of correlated factors typically including income, living situation, and parental formal education. This heterogeneous factor has been discussed repeatedly in the literature on language acquisition, saliently in Hart and Risley (1995), who focused on English-learning American infants, but also in much other work, for instance, that documenting variation among infants growing up in Guatemalan rural households (Klein et al., 1977). There are numerous pathways through

which SES could potentially account for structured variance in input quantity (see Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017; Schwab & Lew-Williams, 2016 for recent reviews). To mention just three, all else equal: (a) families with lower SES may experience harsher living conditions, with negative consequences for their emotional well-being leading to poorer infant–caretaker attachment (e.g., Hackman, Farah, & Meaney, 2010, p. 653 ff.); (b) there may be higher infant and child mortality in low SES settings, in which case it may be adaptive for parents to be less attached to their infant (e.g., Mastin, 2013, p. 171); and (c) parents with higher formal education may be more verbal and/or value to a greater extent verbal achievement (and eventually educational attainment) in their child (e.g., Richman et al., 1992, p. 619). Notice that the former two explanations make general predictions regarding caregiver–child interaction (i.e., frequency of all forms of positively valenced interaction should be reduced in lower compared to higher SES), but not all previous work supports such a broad effect. For instance, Tulkin and Kagan (1972), who studied Boston middle- and low-class families, found the greatest effects on vocalization quantity, and few SES differences in, for example, time spent in face-to-face interaction or within 2 feet of each other. Similarly, in their study of variation among rural Guatemalan families, Klein et al. (1977) conclude that the strongest correlations with SES are with verbal behavior rather than physical proximity. In any case, a great deal of evidence on individual variation within societies suggests that SES correlates with quantity of infant-directed caregiver vocalizations; and it is possible that this factor may also explain some variation between societies.

Extant data suggest that SES may also account for structured variance within and across cultures in terms of developmental changes in quantity of directed speech. By and large, it appears that children get more speech as they age. Older infants may be able to increase their directed verbal input by producing more advanced babbling patterns (Gros-Louis, West, Goldstein, & King, 2006; Warlaumont, Richards, Gilkerson, & Oller, 2014), exhibiting more nonverbal communicative signals (such as pointing, Wu & Gros-Louis, 2015), or simply by ambling toward people who are more likely to interact with them. Insofar as these productive and communicative patterns vary across groups, then we should observe variation in the magnitude of the age-related changes in directed speech frequency. Previous work supports the prediction that infants vary systematically in the amount of vocalizations and/or pointing they spontaneously produce, both in the study of SES-related variation (within industrial and preindustrial samples: Klein et al., 1977; Rowe & Goldin-Meadow, 2009; Warlaumont et al., 2014; but see Eilers et al., 1993) as well as across cultures (Salomo & Liszkowski, 2013). Conceptually, a second source of variation in developmental paths may relate to the way in which interlocutors respond to such early communicative gestures; that is, if a group of parents tends to respond to the child's vocalizations and to do so verbally, then the emergence of vocalizations will cause an increase in infant-directed speech, which will not be apparent if parents do not respond in this fashion. The evidence on this is, at present, mixed (see McGillion et al., 2013; Richman et al., 1992; Warlaumont et al., 2014).

We would like to discuss three factors that are sometimes invoked as explanatory but do not seem to us to be promising structuring factors explaining within- and between-culture variance. The first is household or community size, which is problematic both on conceptual and empirical grounds. As for conceptual pathways, one could propose that infants in smaller families can enjoy proportionally more attention from their caretakers and thus receive more speech. In this sense, number of siblings and interbirth intervals would be relevant variables (similar to comparisons between first-born vs. later born Western infants, e.g., Hoff-Ginsberg, 1998). In fact, small families that are isolated could appear as an ideal language acquisition setting, as the infant does not even need to compete with fellow adults for the maternal attention. However, it is also the case that the more people are present, the larger the number of potential speaking partners—particularly in cultures where young children are free to interact with others. So does presence of siblings, in particular, and others, in general, increase or decrease the quantity of speech addressed to young children when all cultures are taken into account? Although different researchers have studied different parameters (Vogt et al., 2015 provide number of people living in the household, Shneidman & Goldin-Meadow, 2012 the number of people present in the video recordings, and we report number of people in the same location), inspection of previously published results appears to support the idea that more people leads to less directed one-on-one speech: Children in the Mayan and Mozambique settings have averages of 7–8 people (in the video recording and household, respectively) and received between 40 and 240 directed utterances per hour, whereas American

and Dutch children, with about three people in their environment, heard 400–650 sentences. This factor, however, does not account for variance in our data, as similar group sizes are found surrounding children of different ages (averaging 5.9–7.7), who differ greatly in terms of frequency of directed speech. In fact, Konner (1977) argues that there may be psychobiological reasons why mothers may actually interact *less* with their child when there are fewer people around and believes the high caretaker vocalization frequency found among the !Kung may be attributed to the fact that the child–mother dyad come into daily contact with a large number of people.

The second factor whose effect is unclear pertains to who cares for the child. Some anthropological literature describes caretaking as being less mother centered in preindustrial than industrial societies, involving instead a greater investment of the community and particularly older siblings (e.g., Weisner & Gallimore, 1977). This, per se, does not appear to us as a very likely factor explaining variance in overall quantity of speech addressed to the child, unless one further assumes that mothers are more likely to address infants than siblings are. Although the opposite has been reported based on observations of one culture, the Kaluli (Ochs & Schieffelin, 2001), we do not know of any quantitative measurements directly supporting such a statement, and there is one data point contradicting it. As mentioned in the Introduction, Harkness (1977) reports that, among the Kipsigis, a higher proportion of time spent with adults involves speaking than that spent with children. We believe it is worthwhile to withhold judgment until systematic observations are made on matters such as this, as nonsystematic observations are more liable to biases induced, for instance, by the salience of certain behaviors from the culturally and/or theoretically loaded observers' viewpoint (see Johnson, 1975 for further arguments). To give an example from our own work, although older siblings certainly help out more as the infant grows, the mother still remains the primary caregiver among the Tsimane, providing 70% of caregiving in the first 6 years (Winking et al., 2009)—and, as our data reveal, a majority of the spoken input during infancy.

A final factor with limited promise in the search for explanations for variation across cultures is parental reported attitudes to infants' comprehension abilities. Statements about parental attitudes as a potential explanatory factor for (not) talking to infants are common in the literature (see Dixon,

Tronick, Keefer, & Brazelton, 1977, p. 155; Ochs & Schieffelin, 2001, pp. 304–305; Shneidman, 2010, pp. 24–27; Mastin, 2013, chapter 6). This factor is conceptually sound: It makes perfect sense for someone who believes infants do not have a mind to consider talking to an infant nonsensical. We believe, however, that it will be extremely challenging to study such a factor in ways that allow comparisons across cultures. Additionally, there is sometimes a disconnect between explicitly reported beliefs and actual behavior. A good example of a recent study that addresses all of these desiderata is Weber, Fernald, and Diop (2017), who evaluate the effects of an intervention program targeted at changing Senegalese caregivers' beliefs about, and prevalence of, infant-directed speech (see also Johnson & Behrens, 1989 for arguments on the relative independence of beliefs and behavior, and thus the complementarity of phenomenological and observational methods).

Implications for Theories of Early Language Acquisition

Before closing, we would like to add a speculation regarding the relevance of these results to the study of language acquisition. Regardless of the proximal and distal causes behind variation in verbal behavior across cultures, it appears obvious that we cannot assume that cross-cultural differences in the quantity of speech directed to children disappear by age 3 and even less that an increase in spoken input at about 2 years of age compensates for low levels of input early on (see Supporting Information, Section 7 for an attempt to estimate cross-cultural variation). What are the implications of input quantity variation for language acquisition?

Initially, claims regarding cross-cultural differences in quantity and quality of input were considered in the context of discussions on syntactic acquisition and were thus a centerpiece in nativist–emergentist discussions (Lieven, 1994; Pinker, 1995). However, we claim that differences in input quantity are a great deal more relevant for phonetic–phonological and lexical acquisition, where experience must have an equal, if not greater, role than for syntax. As for phonetics and phonology, it is widely agreed upon that children determine the contents of the native sound inventory and, to a more limited extent, more abstract properties of the sound system on the basis of their native exposure within the two years of life (Dupoux, Peperkamp, & Sebastián-Gallés, 2010; Werker & Yeung, 2005). Some describe this aspect of language acquisition as being similar to that found in certain songbird species, for whom the end of the critical period is the

joint result of maturation and exposure to a tutor who produces contingent input (Kuhl, 2004; Werker & Hensch, 2015). But what if a child hears 10 times less directed input than the WEIRD children who are commonly studied? Is the sensitive period "held open" 10 times longer, or is the system for phonetic learning extremely conservative, requiring only minimal levels of exposure? Or are perhaps different learning mechanisms employed in such diverse scenarios?

The same uncertainties emerge in the study of lexical acquisition. Some theorists believe that words can be learned in a "big data" fashion, simply calculating statistics between context (and objects in the context) and words heard (Smith & Yu, 2008)—a process that may require great amounts of data to get rid of spurious correlations. Others argue that tutors can enable children to employ more informed strategies even with relatively low-input quantity, for instance, by producing high-quality learning instances (e.g., speaking the word clearly when an object comes into view and the child is focused on it; Cartmill et al., 2013). If there are few instances of one-on-one conversation, it is likely that not only amount of input is affected, but also that there are fewer chances that the interlocutor follows a child's attention and otherwise ensures high-quality learning instances.

In short, the question of diversity in early language experiences is key also for mechanistic theories of early language acquisition. Future work could employ computational modeling of acquisition to develop more precise predictions regarding the aspects of language most likely to be affected by the large variation in input quantity across cultures. It would be particularly useful to investigate which putative learning mechanisms may be relatively resilient and thus more likely to be cross-culturally relevant.

Limitations

Before concluding, we would like to discuss some limitations to the results presented in this article, and how these may change the ensuing conclusions. Conceptually, we can imagine several sources of underestimation and misestimation in our data. We may underestimate directed and undirected speech because in this time allocation method only two activities are coded at any given time. As noted in the Method section, which two activities are coded depended on the observed person's focus of attention; thus, if talking appeared to the observer to be the third activity in which a participant was engaged (or if speech is integrated into some other activity, such as play), then it will not have been noted by the observer. We attempted a simulation to estimate the maximum impact of this source of underestimation (see Supporting Information, Section 6). In a nutshell, we counted as Directed-F all scans where both activities are filled and a person in the focal group is coded as an interactant in at least one of them. Our interpretation of this simulation is that, if the worst underestimation scenario is true, our estimates would maximally be multiplied by four. This would place the Tsimane results closer to the rural Guatemalan data. Nonetheless, we think that this consideration does not entail that our estimates of purposeful, directed, one-on-one speech, on which we base our main conclusions, are inaccurate. At best, infants and children could be part of a social group where a conversation is occurring or the recipients of speech with an interlocutor who is more engaged with other activities, which remains a poorer setting for learning than the situations often described in industrialized societies (see Lancy, 2007 for arguments that such learning situations are, in fact, cross-culturally rare).

As for misestimation, all observations were carried out during daylight, and thus night-time activities are not included. Night time could contain episodes of speech activity as cluster members gather round the fire or pot (Wiessner, 2014) but also long intervals of silence as they sleep. Future work with 24-hr observations, possibly using a voice-activated recorder, may more accurately estimate precise speech quantities. Such recordings would also allow to go from "amount of time spent talking" to quantities expressed in linguistic units, such as utterances and words. Although we acknowledge the limitations in terms of potential misestimation, the heart of our argument remains on a comparison of speech involving young children versus others present in the same slices as interactants. Therefore, any consideration of misestimation does not challenge our main conclusions as stated. They only make it obvious that any measurement entails some error, and thus one should not assume that these frequency estimates represent literal quantities of speech experienced by the child.

An additional limitation pertains to the composition of our data, which does not allow us to study developmental trajectories. Indeed, we did not employ a clear longitudinal or strict cross-sectional design, given that the same participants "functioned" as focal or nonfocal depending on the analysis; and, given that a cluster was visited several times, the same individual was observed repeatedly

over time within a relatively narrow temporal window. The visits were not arranged systematically to track developmental changes at the individual level, which limits our ability to describe developmental trajectories. Along the same lines, we would look forward to including concurrent language processing measures and/or language outcome measures, so as to be able to assess the strength of the impact that input speech has among children in this community. Such an enterprise may require samples much larger than those we study at present (with *N*s between 9 and 24 per age group).

Open Questions

Several questions remain open for future research, the first one potentially relating to the impact of input on children's language advancement in preindustrial societies. So far, there are divergent findings among the little work looking at the predictive value of quantity of input with respect to vocabulary outcomes (e.g., Shneidman & Goldin-Meadow, 2012 reporting a significant relationship among their Mayan sample which is replicated by Vogt & Mastin, 2013 in their urban sample but not in their rural sample) as well as the importance of speech by secondary caregivers, including children (Harkness, 1977; Mastin & Vogt, 2015; Shneidman et al., 2013).

More generally, it would be theoretically relevant to study not only effects at the level of the individual but also at the level of whole populations. Today, processing-based theories of language acquisition are gaining ground, as they draw support from within-population variation findings showing that children receiving fewer directed verbal interactions are slower to process speech (Weisleder & Fernald, 2013), know fewer words (Deanda, Arias-Trejo, Poulin-Dubois, Zesiger, & Friend, 2016), and produce less complex utterances (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002) among American samples. Indeed, these results are more easily predicted from theories in which "language acquisition is nothing more than learning to process" (Christiansen & Chater, 2016) than theories in which only a critical quantity of evidence is required (Chomsky, 1959; but see Yang, Crain, Berwick, Chomsky, & Bolhuis, 2017). Yet, if the former are universally correct, we should predict striking differences in language outcomes across cultures. If we do not observe such differences across cultures, then this would mean that "acquisition = processing" theories may not generalize to non-WEIRD human populations. We thus end with a strong call for both empirical work evaluating these predictions, and theoretical work exploring alternative accounts considering both stability and diversity in language acquisition.

References

Baker-Henningham, H., & López Boo, F. (2010). *Early childhood stimulation interventions in developing countries: A comprehensive literature review* (Vol. 5282 of the Discussion Paper Series). Bonn, Germany: Institute for the Study of Labor (IZA).

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. *Monographs of the Society for Research in Child Development*, 50(Serial No. 1/2), 3–35. https://doi.org/10.2307/3333824

Campbell, L. (2012). Classification of the indigenous languages of South America. In L. Campbell & V. Grondona (Eds.), *The indigenous languages of South America: A comprehensive guide* (pp. 59–166). Berlin: De Gruyter.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 11278–11283. https://doi.org/10.1073/pnas.1309518110

Chomsky, N. (1959). A review of BF Skinner's verbal behavior. *Language*, 35(1), 26–58.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, e62. https://doi.org/10.1017/S0140525X1500031X

Dahl, D. B. (2009). *xtable: Export tables to latex or html*. R package version, 1.7–0, URL http://CRAN.R-project.org/package=xtable.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, 120, 1956–1972. https://doi.org/10.1016/j.lingua.2010.02.005

Deanda, S., Arias-Trejo, N., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). Minimal second language exposure, SES, and early word comprehension. *Bilingualism: Language and Cognition*, 19, 162–180. https://doi.org/10.1017/s1366728914000820

Dixon, S., Tronick, E., Keefer, C., & Brazelton, T. B. (1977). Mother–infant interaction among the Gusii of Kenya. In P. H. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 149–170). San Diego, CA: Academic Press.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited. *Cognition*, 114, 266–275. https://doi.org/10.1016/j.cognition.2009.10.001

Eilers, R. E., Oller, D. K., Levine, S., Basinger, D., Lynch, M. P., & Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behavior and Development*, 16, 297–315. https://doi.org/10.1016/0163-6383(93)80037-9

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to

infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30, 509–516. https://doi.org/10.1177/0165025406071914

Gross, D. R. (1984). Time allocation: A tool for the study of cultural behavior. *Annual Review of Anthropology*, 13, 519–558. https://doi.org/10.1146/annurev.an.13.100184.002511

Gurven, M., Kaplan, H., & Supa, A. Z. (2007). Mortality experience of Tsimane Amerindians of Bolivia: Regional variation and temporal trends. *American Journal of Human Biology*, 19, 376–398. https://doi.org/10.1002/ajhb.20600

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience*, 11, 651–659. https://doi.org/10.1038/nrn2897

Harkness, S. (1977). Aspects of social environment and first language acquisition in rural Africa. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 309–316). Cambridge: Cambridge University Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29. https://doi.org/10.1038/466029a

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, 26, 1071–1083. https://doi.org/10.1177/0956797615581493

Hodson, H. (2014). Automatic voice coach gives conversation tips to parents. *New Scientist*, 221, 22. https://doi.org/10.1016/S0262-4079(14)60226-8

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73, 418–433. https://doi.org/10.1017/S0305000907008343

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19, 603–629.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45, 337–374.

Johnson, A. (1975). Time allocation in a Machiguenga community. *Ethnology*, 14, 301–310. https://doi.org/10.2307/3773258

Johnson, A., & Behrens, C. (1989). Time allocation research and aspects of method in cross-cultural comparison. *Journal of Quantitative Anthropology*, 1, 234–245.

Kaplan, H., Hooper, P. L., Stieglitz, J., & Gurven, M. (2015). The causal relationship between fertility and infant mortality. In P. Kreager, B. Winney, S. Ulijaszek, & C. Capelli (Eds.), *Population in the human sciences: Concepts, models, evidence* (pp. 361–376). Oxford: Oxford University Press.

Klein, R. E., Lasky, R. E., Yarbrough, C., Habicht, J., & Sellers, M. J. (1977). Relationship of infant/caretaker interaction, social class and nutritional status to developmental test performance among Guatemalan infants. In P. H. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 385–403). San Diego, CA: Academic Press.

Konner, M. (1977). Infancy among the Kalahari desert San. In P. H. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 287–328). San Diego, CA: Academic Press.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843. https://doi.org/10.1038/nrn1533

Lancy, D. F. (2007). Accounting for variability in mother-child play. *American Anthropologist*, 109, 273–284. https://doi.org/10.1525/aa.2007.109.2.273

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge: Cambridge University Press.

Mastin, J. D. (2013). *Exploring infant engagement, language socialization and vocabulary development*. Unpublished doctoral dissertation, Tilburg University.

Mastin, J. D., & Vogt, P. (2015). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, 235–264. https://doi.org/10.1017/S0305000915000148

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development*, 5, 240–248.

Mulder, M. B., Caro, T. M., Chrisholm, J. S., Dumont, J.-P., Hall, R. L., Hinde, R. A., & Ohtsuka, R. (1985). The use of quantitative observational techniques in anthropology. *Current Anthropology*, 26, 323–335. https://doi.org/10.1086/203277

Ochs, E., & Schieffelin, B. (1995). Language acquisition and socialization: Three developmental stories and their implications. In B. G. Blount (Ed.), *Language, Culture, and Society* (pp. 470–512). Illinois: Waveland Press.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, 285–308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Pinker, S. (1994). *The language instinct: The new science of language and mind*. New York: William Morrow and Co.

R Core Team. (2015). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

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

Richman, A. L., Miller, P. M., & LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology*, *28*, 614–621. https://doi.org/10.1037/0012-1649.28.4.614

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*, 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*, 1762–1774. https://doi.org/10.1126/science.1167025

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*, 951–953. https://doi.org/10.1126/science.1167025

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Development*, *84*, 1296–1307. https://doi.org/10.1111/cdev.12026

Schwab, J. F., & Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*, 264–275. https://doi.org/10.1002/wcs.1393

Sellen, D. W., & Smay, D. B. (2001). Relationship between subsistence and age at weaning in "preindustrial" societies. *Human Nature*, *12*, 47–87.

Shneidman, L. A. (2010). *Language input and acquisition in a Mayan village*. Unpublished doctoral dissertation, The University of Chicago.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*, 672–686. https://doi.org/10.1017/S0305000912000141

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*, 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Stieglitz, J., Beheim, B. A., Trumble, B. C., Madimenos, F. C., Kaplan, H., & Gurven, M. (2015). Low mineral density of a weight-bearing bone among adult women in a high fertility population. *American Journal of Physical Anthropology*, *156*, 637–648. https://doi.org/10.1002/ajpa.22681

Stieglitz, J., Gurven, M., Kaplan, H., & Hooper, P. L. (2013). Household task delegation among high-fertility forager-horticulturalists of Lowland Bolivia. *Current Anthropology*, *54*, 232–241. https://doi.org/10.1086/669708

Street, R. L., Jr., & Cappella, J. (1989). Social and linguistic factors influencing adaptation in children's speech. *Journal of Psycholinguistic Research*, *18*, 497–519.

Tulkin, S. R. (1977). Social class differences in maternal and infant behavior. In P. H. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 495–537). San Diego, CA: Academic Press.

Tulkin, S. R., & Kagan, J. (1972). Mother–child interaction in the first year of life. *Child Development*, *43*, 31–41. https://doi.org/10.2307/1127869

Vogt, P., & Mastin, J. D. (2013). Rural and urban differences in language socialization and early vocabulary development in Mozambique. In M. Knauf, M. Pauen, M. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 3787–3792. Austin, TX: Cognitive Science Society

Vogt, P., Mastin, J. D., & Schots, D. M. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands, and rural and urban Mozambique. *First Language*, *35*, 341–358. https://doi.org/10.1177/0142723715596647

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*, 1314–1324. https://doi.org/10.1177/0956797614531023

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*. *88*, 1513–1526 https://doi.org/10.1111/cdev.12882.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*, 2143–2152. https://doi.org/10.1177/0956797613488145

Weisner, T. S., & Gallimore, R. (1977). My brother's keeper: Child and sibling caretaking. *Current Anthropology*, *18*, 169–190. https://doi.org/10.1086/201883

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*, 519–527. https://doi.org/10.1016/j.tics.2005.09.003

Wiessner, P. W. (2014). Embers of society: Firelight talk among the Ju/'Hoansi Bushmen. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 14027–14035. https://doi.org/10.1073/pnas.1404212111

Winking, J., Gurven, M., Kaplan, H., & Stieglitz, J. (2009). The goals of direct paternal care among a South Amerindian population. *American Journal of Physical Anthropology*, *139*, 295–304. https://doi.org/10.1002/ajpa.20981

Wu, Z., & Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language*, *42*, 538–561. https://doi.org/10.1017/S0305000914000221

Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. *Implementing Reproducible Research*, *1*, 20.

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews* https://doi.org/10.1016/j.neubiorev.2016.12.023.