AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0001688

Mealtime Conversations Between Parents and Their 2-Year-Old Children in Five Cultural Contexts

Manuel Bohn[1, 2], Wilson Filipe da Silva Vieira[2], Marta Giner Torréns[3], Joscha Kärtner[3], Shoji Itakura[4], Lília Cavalcante[5], Daniel Haun[2], Moritz Köster[6], and Patricia Kanngiesser[7]

[1] Institute of Psychology in Education, Leuphana University Lüneburg
[2] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[3] Department of Psychology, University of Münster
[4] Center for Baby Science, Doshisha University
[5] Graduate Program in Theory and Research of Behavior, Federal University of Pará
[6] Institute of Psychology, University of Regensburg
[7] School of Psychology, University of Plymouth

Children all over the world learn language, yet the contexts in which they do so vary substantially. This variation needs to be systematically quantified to build robust and generalizable theories of language acquisition. We compared communicative interactions between parents and their 2-year-old children ($N = 99$ families) during mealtime across five cultural settings (Brazil, Ecuador, Argentina, Germany, and Japan) and coded the amount of talk and gestures as well as their conversational embedding (interlocutors, function, and themes). We found a comparable pattern of communicative interactions across cultural settings, which were modified in ways that are consistent with local norms and values. These results suggest that children encounter similarly structured communicative environments across diverse cultural contexts and will inform theories of language learning.

Public Significance Statement

Cultural norms and beliefs structure social interactions and communication. As a consequence, children learn language under very different circumstances. We studied communicative interactions between parents and their children in five diverse cultural contexts. We found a common, child-centered pattern of communication that was modified in line with local norms and values. This suggests that children can rely on similar information sources and learning processes across cultural contexts.

Keywords: language acquisition, communication, gesture, cross-cultural psychology, parent–child interaction

Children learn language in interactions with language-competent others (Bohn & Frank, 2019; Bruner, 1983; Clark, 2009; Levinson & Holler, 2014; Tomasello, 2008). Social interactions between children and their social partners are structured by norms, values, and beliefs that vary substantially across cultural and historical contexts (Rogoff, 2003). As a consequence, children may encounter dramatically different language learning environments. Yet, the fact that children usually achieve fluency in their local language(s) suggests that they use a suite of compensatory learning strategies to adapt flexibly to their respective learning environment (Cristia, 2022; Kidd &

Manuel Bohn https://orcid.org/0000-0001-6006-1348
Moritz Köster and Patricia Kanngiesser are joint senior authors. This research was funded by a grant from Volkswagen Foundation (89611-2), awarded to Patricia Kanngiesser. The data assessment in Kyoto was partly funded by a Japan Society for the Promotion of Science Fellowship to Moritz Köster. Manuel Bohn was supported by a Jacobs Foundation Research Fellowship (2022-1484-00). The authors would like to thank all the research assistants who supported the data assessments and coding: Phileas Heim, Chisato Fukuda, Julia Ohlendorf, and Marlene Abromeit. The authors are grateful to Luke Maurits for statistical advice. All data and analysis code can be found in the following repository: https://github.com/ccp-eva/mealtime.

Manuel Bohn served as lead for conceptualization, formal analysis, methodology, visualization, writing–original draft, and writing–review and editing. Wilson Filipe da Silva Vieira, Marta Giner Torréns, Joscha Kärtner, Shoji Itakura, Lília Cavalcante, Daniel Haun, Moritz Köster, and Patricia Kanngiesser contributed equally to investigation and writing–review and editing. Marta Giner Torréns, Joscha Kärtner, Moritz Köster, and Patricia Kanngiesser contributed equally to conceptualization and methodology.

Correspondence concerning this article should be addressed to Manuel Bohn, Institute for Psychology in Education, Leuphana University Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany. Email: manuel_bohn@leuphana.de

Garcia, 2022; Rowe & Weisleder, 2020). Explaining how children accomplish this feat poses a serious theoretical and empirical challenge. Detailed documentation of learning environments across cultural contexts is needed to inform theorizing about children's learning processes. In this article, we contribute to this effort by reporting on cross-cultural variation in parent–child communicative interactions in a semi-structured setting: meals involving parents and their 2-year-old child.

In recent decades, research on language acquisition has focused, to a large extent, on variation in language input and, in particular, the number of words children hear in naturalistic settings. This line of work was sparked by the finding that children who receive more input—especially speech directly addressing them—have larger vocabularies (Bang, Bohn, et al., 2023; Hart & Risley, 1995; Huttenlocher et al., 1991; Shneidman & Goldin-Meadow, 2012; Walker et al., 1994; Weisleder & Fernald, 2013). From a theoretical perspective, more language input increases children's opportunities for learning word-meaning mappings and allows them to build a larger vocabulary (Jones & Rowland, 2017; Kachergis et al., 2022; McMurray et al., 2012). The introduction of daylong audio recording devices and automated coding algorithms has provided further momentum to this endeavor (Cristia et al., 2021; Greenwood et al., 2011; Lavechin et al., 2020). As a consequence, the quantity of direct language input plays a central role in theories and formal models of language learning (Braginsky et al., 2019; Goodman et al., 2008; Kachergis et al., 2022; Swingley & Humphrey, 2018).

However, like most developmental psychology (Amir & McAuliffe, 2020; Nielsen et al., 2017), research on language acquisition has largely focused on affluent societies of the Global North, and the resulting theoretical proposals may fail to generalize to other cultural contexts. As studies in a greater variety of cultural settings have begun to accumulate (Altınkamış et al., 2014; Bergelson et al., 2019; Bunce et al., 2020; Casillas et al., 2021; Choi, 2000; Cristia et al., 2019; Loukatou et al., 2021; Tardif et al., 1997), they have revealed substantial cultural variation in how much direct input children receive (Cristia, 2022; see also Sperry et al., 2019 for variation within an English-speaking sample). Yet, children still reach major milestones in language development at similar ages (Brown & Gaskins, 2014; Casillas et al., 2020). These findings highlight that theories and models of language learning need to extend beyond the quantity of input and also include learning processes that compensate for variation in input (Bang, Mora, et al., 2022; Casillas, 2022; Jones & Rowland, 2017; Kachergis et al., 2022; Meylan & Bergelson, 2022).

It has been suggested that these compensatory learning processes leverage structural features of social interactions in which language is used (Casillas et al., 2020; Rogoff et al., 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). Pragmatic accounts of language learning offer an explanation for how children use contextual information (e.g., Bohn & Frank, 2019; Tomasello, 2008): Social interactions, especially routines, follow predictable patterns that make it easier for children to infer what speakers are communicating about (Barbaro & Fausey, 2022; Bruner, 1983; Lieven, 1994; Masek et al., 2021; Vygotsky, 1978). For instance, Roy et al. (2015) found that words were more easily learned when they were primarily used in a distinct spatial and temporal context. Similarly, establishing common ground over the course of an interaction provides information about the speaker's intention independent of the words that are being used (Bohn & Köymen, 2018; Bohn, Tessler, et al., 2021). For example, Bohn, Le, et al. (2021) showed that children identify the referent of an ambiguous word by inferring the topic of an ongoing conversation (see also Akhtar, 2002). These findings help to explain why the amount of conversational turn-taking in parent–child interactions predicts child language outcomes (Donnelly & Kidd, 2021; Romeo et al., 2018). Turn-taking results in continuous, structured conversations that provide information-rich learning opportunities.

To assess whether children can use structural features to complement direct verbal input, it is crucial to compare communicative interactions between adults and children across cultural settings. However, to our knowledge, there are very few quantitative comparisons. While ethnographic descriptions offer important and rich insights into individual cultural settings (see e.g., De León, 2011; Gaskins, 2006), quantitative comparisons are essential for understanding gradual cultural differences (Broesch et al., 2021; Hewlett et al., 1998; Köster et al., 2022) and offer core input for theory building.

One of the challenges of cross-cultural work lies in selecting an appropriate context for comparing the structure of communicative interactions (Broesch et al., 2022). Prior work has shown that the amount of language input children receive varies substantially across routine activities. For example, Soderstrom and Wittebolle (2013) found that Canadian adults spoke most during book reading and structured playtime (see also Tamis-LeMonda et al., 2019). Such activities, however, are very specific to industrialized societies and less frequent or absent in other cultural contexts. A cross-culturally recurrent, and hence particularly promising, context for cross-cultural research is mealtimes: across societies, meals are social events that are structured by—and used to transmit—cultural norms, values, and beliefs (Blum-Kulka, 2012; Fjellström, 2004; Köster et al., 2022; Ochs & Shohet, 2006). Furthermore, mealtimes have proven fruitful for studying caregiver–child communication in cultural contexts like the United States (e.g., Beals, 1993, 1997; Snow & Beals, 2006).

The Current Study

The goal of this study was to compare communicative interactions between parents and their children during mealtimes across diverse cultural settings. We aimed for a naturalistic but comparable setup by (a) asking families to record in their homes, (b) recruiting families with a single—usually the first—child between 2 and 3 years of age, and (c) focusing on 10-min-long episodes during which three family members (mother, father, one child) were present. Even though the constellation of two parents and one child might be less representative of the overall family demographics in some settings, it allowed us to directly quantify and compare communicative interactions.

We obtained recordings from five different cultural settings, including families living in the Global South and the Global North, as well as in urban and rural settings: the city of Buenos Aires in Argentina, small villages in the Amazon region near Apeú in Brazil), small villages close to Cotacachi in Ecuador, the city of Münster in Germany, and the city of Kyoto in Japan. This sample was, first and foremost, a convenience sample of families in diverse cultural settings we had worked with previously. This continues to be a common approach in larger-scale cross-cultural, developmental studies (House et al., 2020, 2013; Kanngiesser et al., 2022; P. R. Blake et al., 2015) and is often the first step when little substantive cross-cultural data exists to inform targeted comparisons. Nevertheless, in addition to their geographic spread and variation in population density, the settings also varied in cultural norms and beliefs about communication during mealtimes. In Germany, meals are seen as a privileged time for communication and exchange (Danesi, 2018). Similarly, in Argentina, dinners are an important opportunity for family conversations because it is usually the only time when the whole family gets together (Aguirre, 2016). In contrast, within the Kichwa indigenous people in Ecuador, meals are supposed to be taken in silence (Sánchez-Parga, 2010). In Japan, both views are common and whether or not talk is encouraged depends, in part, on the eating arrangements (Imada & Furumitsu, 2020). As such, our sample provided us with the opportunity to study if and how different cultural mealtime norms impact real-world communicative interactions.

We coded and analyzed our video data along several dimensions, focusing on the quantity of talk and gestures as well as their conversational embedding. We chose dimensions that have been implicated as relevant for child language acquisition, but have rarely been studied from a cross-cultural perspective. First, we coded the presence (or absence) of speech, and the identity of the speaker and the recipient. This allowed us to quantify how much directed talk—as opposed to overheard talk—children received and from whom. As noted above, cross-cultural variation in talk directed at the child has profound theoretical implications because it questions the privileged role given to direct input in many theoretical accounts of language learning. Coding speaker identity provided insight into who children receive language input from. Cross-cultural research on different sources of language input is relatively scarce: most past studies have exclusively focused on maternal talk, and only recently have researchers begun to investigate paternal talk (Ferjan Ramírez, 2022). By coding the language produced by children themselves, we were able to quantify children's role in shaping their linguistic environment across cultural contexts (Donnellan et al., 2020; Tamis-LeMonda et al., 2018). In addition to speech, we also coded the production of gestures. A substantial body of research has shown that gestures produced by children and their caregivers relate to child language competency—at least in children growing up in the Global North (Colonnesi et al., 2010; e.g., Rowe et al., 2008). Here, the view is that gestures act as a complementary source of input that references objects and events in the environment and thereby facilitate word learning (Tomasello, 2005).

Second, we coded how utterances were grouped into themes. This approach allowed us to quantify cross-cultural variation in how conversations are structured. Research on conversational turn-taking has suggested a link between these structural features and language learning (Donnelly & Kidd, 2021; Romeo et al., 2018); yet, a cross-cultural perspective is still largely missing. Finally, we coded the function of utterances and distinguished between questions, assertions, and imperatives. Questions play a role in facilitating language acquisition because they encourage verbal responses from children which may include labels for objects (J. Blake et al., 2006). There is also suggestive evidence of cultural variation in how parents use functional elements of language, such as questions (Kuchirko et al., 2020).

For the analysis, we first assessed if and how these coded dimensions differed in the five cultural settings. In a second step, we asked whether some cultural settings are more similar to one another. The five cultural settings offer an interesting perspective on the factors influencing mealtime conversations. For example, communicative interaction patterns could cluster by country (five clusters; one cluster per country) or by language family and geographical region (three clusters; Argentina, Brazil, Ecuador vs. Germany vs. Japan) or by degree of urbanization (two clusters; urban: Argentina, Germany, Japan vs. rural: Brazil, Ecuador). Based on previous work, we expected less direct input to children in rural contexts (Cristia, 2022). Due to different cultural norms around mealtime conversations, we predicted less overall talk in Ecuador compared to Germany, with Japan falling somewhere in the middle. Given a lack of comparable previous work—we had no specific predictions for variation in the structure of communicative interactions.

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data and analysis code can be found in the following repository: https://github.com/ccp-eva/mealtime. Data were analyzed using R, Version 4.2.0 (R Core Team, 2022) and the function brm from the package brms (Bürkner, 2017). We used default priors built into brms for all parameters. The study's design and its analysis were not preregistered.

Participants

The final sample consisted of 99 families from five cultural contexts. This included 20 families from the city of Buenos Aires, Argentina (urban setting), 18 families from villages in the Amazon region near Apeú, Brazil (rural setting), 13 from villages near Cotacachi, Ecuador (rural setting), 24 families from the city of Münster, Germany (urban setting), and 24 families from the city of Kyoto, Japan (urban setting). For the recording sessions, all families comprised a father, a mother, and a child aged between 2 and 3 years, 2 months. Almost all children were the first child in the family. Some videos partly included additional children ($n = 1$ for Argentina, Brazil, and Ecuador, respectively).

Additional families were recorded, but they did not meet the inclusion criteria of at least one recording of a meal that lasted for at least 10 min, initially included all three family members and had all family members visible in the recording. This resulted in the exclusion of 11 families from Münster, Germany; 34 from Apeú, Brazil; five from Buenos Aires, Argentina; 39 from Cotacachi, Ecuador; and five from Kyoto, Japan.

The recordings were collected as part of a larger cross-cultural investigation into parent–child interactions, and findings on parental teaching behaviors have been published by Köster et al. (2022). We refer to this earlier work for a detailed description of each cultural setting. In the following, we only provide a short overview.

Argentina

Families lived in the metropolitan area of Buenos Aires, Argentina, which comprises around 15.2 million people. They were recruited via personal contacts of the local experimenter. The family language was Rioplatense Spanish. Compensation included small toys for children and USD 10 for parents. Most parents had completed a university degree (mothers: 74%; fathers: 52%) and engaged in paid professional labor (mothers: 87%; fathers: 78%). The majority of children (91%) either attended kindergarten or were looked after by a nanny or a family member other than the parents.

Brazil

Families lived in villages of around 50–300 families in the Amazon region near Apeú, approximately 1.5 hr east of Belém, the capital of the state of Pará. They were recruited with the help of a local public health office. The family language was Brazilian Portuguese. Compensation included small toys for children and a certificate of participation for parents. Most parents had completed secondary school (~12 years of schooling, mothers: 50%; fathers: 56%). Mothers worked mainly as housewives (83%), while fathers engaged in paid labor (100%). Some families engaged in traditional subsistence activities such as tapioca farming, livestock breeding, or açaí and fruit harvesting. In line with employment status, the majority of children were looked after by their mothers.

Ecuador

Families self-identified as belonging to the Kichwa community and lived in villages with 800–5,000 inhabitants located within 1 hr (by car) of the city of Cotacachi in the Imbabura province. They were recruited via personal contacts mediated by the community president. The family language was Ecuadorian Spanish with elements of Kichwa. Compensation included food (e.g., rice or oats) and USD 4. Most parents had completed primary school (~10 years of schooling, mothers: 50%; fathers: 56%). Mothers worked mainly as housewives (59%), while fathers engaged in paid labor (77%). Around 40% of children were looked after by a person other than the mother during the day.

Germany

Families lived in Münster in the state of North-Rhine-Westphalia, a city with ~310,000 inhabitants. They were recruited via a participant database of the Developmental Psychology lab at the University of Münster. Compensation included a voucher of EUR 15 for a local toy store. Most parents had completed a university degree (mothers: 71%; fathers: 71%) and engaged in paid professional labor (mothers: 92%; fathers: 92%). All children either attended kindergarten or were looked after by a nanny during the day.

Japan

Families lived in the city of Kyoto, in the Kansai metropolitan region, with around 1.5 million inhabitants. They were recruited via a participant database of the Center for Baby Science at Doshisha University. Compensation was JPY 3000. Most parents had completed a university degree (mothers: 92%; fathers: 83%) and engaged in paid professional labor (mothers: 71%; fathers: 100%). Most children (80%) attended kindergarten.

The study was approved by the ethics committee of the Free University of Berlin. Recordings took place between September 2017 and March 2019. Informed verbal consent was obtained from both parents and written consent from one of the parents.

Procedure

We visited families twice. On the first visit, an experimenter (familiar with the local language) instructed parents on how to use the video camera and what to record. We encouraged families to record two instances of the meal they commonly shared together, which happened in the evening for most families. The cameras were equipped with a wide-angle lens and set up to capture all family members during the meal. In addition to video, the cameras also recorded sound. On the second visit, the experimenter asked about the recordings and encouraged families to record additional meals if they had not already recorded two sessions. In the end, we collected sociodemographic information and interviewed the mothers (unrelated to the present study).

Coding

We scanned all recordings for sections that captured a meal event, lasted at least 10 min, and included all three family members. For each family, we selected one such section for in-depth coding and excluded all families for which we did not find such a section (see above for the number of excluded families).

We coded videos using ELAN (Wittenburg et al., 2006) Version 6.4. The primary coder was either a native (Germany, Japan, Brazil) or a highly fluent (Argentina, Ecuador) speaker of the local language. For Ecuador, a native speaker translated sections containing Kichwa into Spanish before the primary coder coded them.

In the first pass, the primary coder created a tier for each speaker and marked segments in which this person was speaking or using a gesture. In a second pass, the coder transcribed all utterances into the local language and coded their conversational embedding. We defined utterances as sections of continuous talk by one person. If speakers paused for more than 2 s, we coded two utterances with 2 (or more) seconds of silence in between. We used the following codes to capture the conversational embedding of each utterance:

Speaker

Here, we coded who produced the utterance. The speaker could either be the child, the mother, or the father. All sections containing no speech were coded as non-talk.

Recipient

Here, we coded who the utterance was addressed to. Codes could either be child, mother, father, both, or other, where other was used either when a fourth person (e.g., over the phone) was addressed, or the speaker was talking to themselves (e.g., child babbling or singing). If an utterance addressed two people in sequence, the second addressee was coded as the recipient.

Themes and Utterances

Here, we coded the conversational coherence of the different utterances. For that, we defined themes as sequences of utterances that related to one another. This applies, for example, to sequences of questions and answers but also to sequences in which the content of an utterance is directly related to the content of the previous utterance. Please note that such themes were coded locally and were not the same as topics. For example, if the father and child exchanged four utterances about the child's day in the kindergarten this was coded as one theme. If the same topic (day at the kindergarten) came up later again, this was coded as a separate theme. Each utterance within a theme was counted to capture the sequence and length of a theme. Thus, each utterance was assigned a number for the theme and a number for the utterance within the theme. Themes could have interjections of one or two utterances. After more than two interjections, we coded a new theme. For example, if the father and child talked about food and the mother made an unrelated comment in between, the mother's comment would be coded as a separate theme while the other theme continued around it:

Child:	"I want more" (theme [t] 1, utterance [u] 1)
Father:	"Do you want more soup?" (t1, u2)
Mother:	"Phew, I'm hot (t2, u1)
Child:	"No, bread (t1, u3)
Father:	"I'll get some" (t1, u4)

Functional Elements

Each utterance was coded as either being a question, assertion, or imperative. Imperatives were only coded if the utterance was grammatically structured as an imperative. For example, "Pass me the salt!" was coded as an imperative, while "You should give me the salt." was not.

Referential Gestures

We also coded the frequency of two types of referential gestures for each individual. Points were coded when someone indicated an object, location, or person in the environment, either using a finger (often index finger), the head, or an object (e.g., cutlery). Reaches and hold-outs were not coded as points. Iconic gestures were coded when someone depicted an object or action using their hands and/or body (e.g., pretending to hold a knife and cut to instruct the child how to cut a cucumber). Conventional gestures such as head shaking, nodding, or shrugging were not coded.

Reliability Coding

For each cultural setting, we selected 15% of videos and had them recoded by a second coder (native speaker of the respective language). The second coder relied on the sequencing of the primary coder. Interrater reliability was generally very good. For the recipient, the agreement between coders was 88% ($\kappa = .83$), for function it was 91% ($\kappa = .78$) and for gestures it was 96% ($\kappa = .81$). To get interrater reliability for the coding of themes, we asked whether the two coders agreed on whether a given utterance belonged to the same theme as the previous utterance or belonged to a new theme. Once again, agreement between coders was high (agreement = 87%, $\kappa = .74$).

Analysis and Results

For each of the research questions (see below), we defined a response variable and then used Bayesian multilevel regression models to model the effect of cultural setting and—whenever applicable—that of the different individuals involved in the conversation. To make inferences about the importance of predictors, we compared a set of nested models including cultural setting and individual as predictors to each other and to a null model that did not include them to test if these predictors improved model fit. Following McElreath (2018), we compared models using Widely Applicable Information Criteria (WAIC). This approach favors models that have high out-of-sample predictive accuracy in that they achieve a good fit to the data with a minimal set of parameters.

We modeled the effect of cultural settings as random effects and interactions between additional variables (e.g., speaker identity) and setting as random slopes within cultural setting, brms notation: (variable|setting). This approach partially pools model estimates and is thought to yield more generalizable results because it avoids overfitting the model to the observed data (Gelman & Hill, 2006; McElreath, 2018). For each model comparison, we report the difference in WAIC estimates, the standard error of the difference, and the weight of each model. Model weights give the probability that a model will make the best predictions out of all the models considered.

Each model comparison has a "winning" model, that is, a model that has the lowest WAIC value and the highest weight and thus, the highest expected out-of-sample predictive accuracy. However, two models can be more or less equivalent when the difference in WAIC is small and the standard error of the difference is larger than the difference in WAIC. In addition to the model comparison, we visualize the predictions of the winning model and interpret them based on their posterior means and 95% credible intervals (CrI).

How Much Time Did Families Spend Talking?

First, we ask how much time families spent talking as opposed to not talking and how this varied across cultural settings. The dependent variable in this case was the total lengths of all sections coded as nontalk for each family (modeled as a normal distribution). We compared a null model including only an overall intercept `(non-talk ~ 1)` to a model including cultural setting, `non-talk ~ 1 + (1|setting)`.[1]

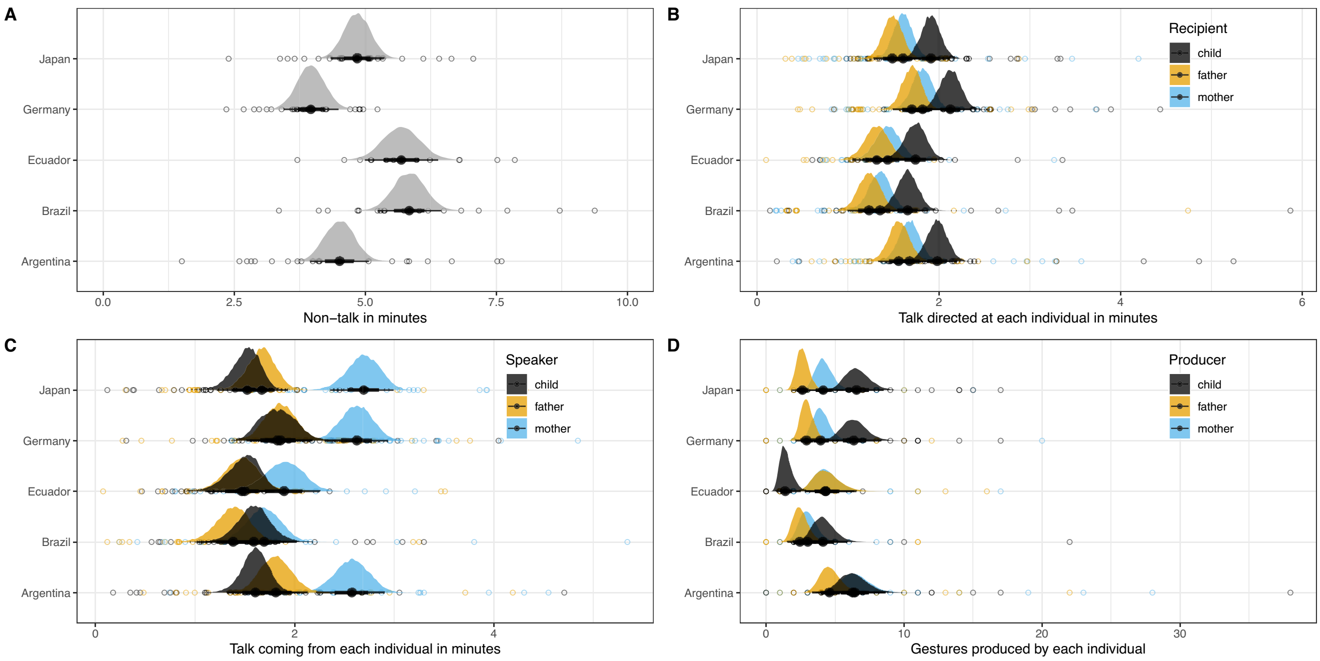
The model comparison clearly favored the model including cultural setting (WAIC = 338.84, $SE = 14.93$, weight > 0.99) over the null model (WAIC = 362.36, $SE = 14.97$, weight < 0.01). The difference in WAIC (dWAIC) was $= -11.76$ with a standard error of 4.15. The model predicted an average of 4.95 95% CrI [3.80–6.07] minutes of nontalk across cultural settings. Ecuador and Brazil had longer sections of nontalk compared to Argentina and Germany, with Japan falling in the middle (see Figure 1A).

How Much Talk Is Directed at Each Family Member?

Next, we asked whom the talk was directed to, that is, how much "input" each family member received. The dependent variable was the total length of utterances directed at each individual in a family. This variable was right-skewed and we therefore modeled it as a

[1] One might suspect that the child's age influences their own behavior or that of the parents. To explore this possibility, we added models including child age as a predictor to the model comparison for the first three models (overall talk, talk per speaker, and talk received by each individual). The inclusion of age did not improve the fit of the otherwise best-fitting model any further. In the interest of space and readability, we do not report these models here. However, age is included in the data set available in the associated repository so interested readers can further explore the relation between age and the variables we coded.

Figure 1
Talk and Gestures



Note. (A) Nontalk across cultural settings. (B) Talk directed at the different individuals. (C) Time spent talking by the different individuals. (D) Number of gestures (points and iconic gestures combined) produced by each individual. In B–D: Shading denotes the individual. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family and—whenever applicable—individual. CrI = credible intervals. See the online article for the color version of this figure.

skewed normal distribution. Given that the analysis above showed that the amount of overall talk differed across cultural settings, the null model already included a random effect for setting, `input ~ 1 + (1|setting) + (1|family)`. We compared it to two alternative models, one assuming that input additionally differed across recipients, `input ~ recipient + (1|setting) + (1|family)`, and one assuming that this effect, in turn, varies across settings, `input ~ recipient + (recipient|setting) + (1|family)`.

The model comparison clearly favored the two alternative models, with a slight preference for the simpler model that did not assume the effect of recipients to vary across cultural settings; WAIC = 705.72, *SE* = 30.16, weight = 0.74; model assuming variation across settings: WAIC = 707.82, *SE* = 30.15, weight = 0.26; dWAIC = −1.05, *SE* (dWAIC) = 0.85. We observed that, across settings, more talk was directed at children compared to the two parents with fathers being talked to the least (see Figure 1B).

Which Family Member Talks the Most?

In the next analysis, we asked how talking time was distributed across the different family members. The dependent variable was the total lengths of utterances of each individual in a family, which was also right-skewed and modeled as a skewed normal distribution. Given previous results, the null model included a random effect for setting, `talk ~ 1 + (1|setting) + (1|family)`. The first alternative model assumed that talk differed across speakers, `talk ~ recipient + (1|setting) + (1|family)`, the second assumed that this effect interacted with setting, `talk ~ recipient + (recipient|setting) + (1|family)`.

The model comparison clearly favored the interaction model assuming that the difference between speakers varied across settings (WAIC = 755.92, *SE* = 25.20, weight > 0.99; model assuming no interaction: WAIC = 772.14, *SE* = 24.65, weight < 0.01; dWAIC = −8.11, *SE*(dWAIC) = 3.65). Figure 1C shows that even though mothers talked the most in all settings, this effect was much more pronounced in Japan, Germany, and Argentina compared to Ecuador and Brazil.

How Many Gestures Are Being Used?

To conclude the first set of analysis, we looked at variation in gesture production. Iconic gestures were produced at a much lower rate (15%) compared to pointing gestures (85%). Thus, many individuals from different cultural settings did not produce any iconic gestures. This made it difficult to analyze points and iconic gestures separately and we instead decided to combine them. Thus, the dependent variable was the number of gestures produced by each individual. We modeled this distribution as a zero-inflated Poisson distribution to account for the fact that some individuals did not produce any gestures.

The null model only included an intercept and a random effect of family, `gestures ~ 1 + (1|family)`. There were three alternative models: the first included producer (child, mother, father) as a fixed effect, `gestures ~ producer + (1|family)`, the second model added to this a random effect for setting, `gestures ~ producer + (1|setting) + (1|family)` and the third model included an additional random slope for interlocutors within setting to model the interaction, `gestures ~ producer + (producer|setting) + (1|family)`.

The model comparison clearly favored the model assuming that the number of gestures produced varied between individuals within cultural settings—interaction model; WAIC = 1,602.79, *SE* = 49.79, weight > 0.99; second best model (without interaction): WAIC = 1,670.90, *SE* = 53.44, weight < 0.01; dWAIC = −34.06, *SE*(dWAIC) = 14.33. Overall, there were slightly fewer gestures in Ecuador and Brazil. Looking at the different individuals, we saw that—across settings—children produced the most gestures, followed by mothers and then fathers. This pattern was less pronounced in Brazil and Argentina and notably reversed in Ecuador, where children produced hardly any gestures (see Figure 1D).

Who Talks to Whom?

To address the question of who talks to whom we categorized the conversational partners of each utterance as either being mother and father, child and mother, or child and father. We then used a categorical model to predict the proportion with which each of these categories occurred. The null model only included an intercept and a random effect of family, `partners ~ 1 + (1|family)`, while the alternative model assumed that these proportions differ across settings, `partners ~ 1 + (1|setting) + (1|family)`.

The model comparison yielded no clear difference between models, suggesting no substantial differences in the proportion of conversational partners across settings—alternative model: WAIC = 28,106.33, *SE* = 116.90, weight = 0.62; null model: WAIC = 28,107.31, *SE* = 116.83, weight = 0.38; dWAIC = −0.49, *SE*(dWAIC) = 0.80. Compared to an equal split (proportion of 0.33 for each category), conversations between mother and child were slightly more frequent and conversations between child and father less frequent except for Brazil where conversations between mother and father were less likely (see Figure 2A).

Who Uses Which Functional Elements?

As the next step, we analyzed how the different speakers used different functional elements—assertions, imperatives, and questions. That is, we predicted the proportion with which each functional element occurred using a categorical model. We investigated whether the types of functional elements used varied with speakers as well as cultural settings. The null model only included an intercept and a random effect of family, `function ~ 1 + (1|family)`. There were three alternative models: the first included speaker as an additional fixed effect, `function ~ speaker + (1|family)`, the second model added to this a random effect for setting, `function ~ speaker + (1|setting) + (1|family)`, and the third model included and additional random slope for speaker within setting to model the interaction between speaker and setting, `function ~ speaker + (speaker|setting) + (1|family)`.

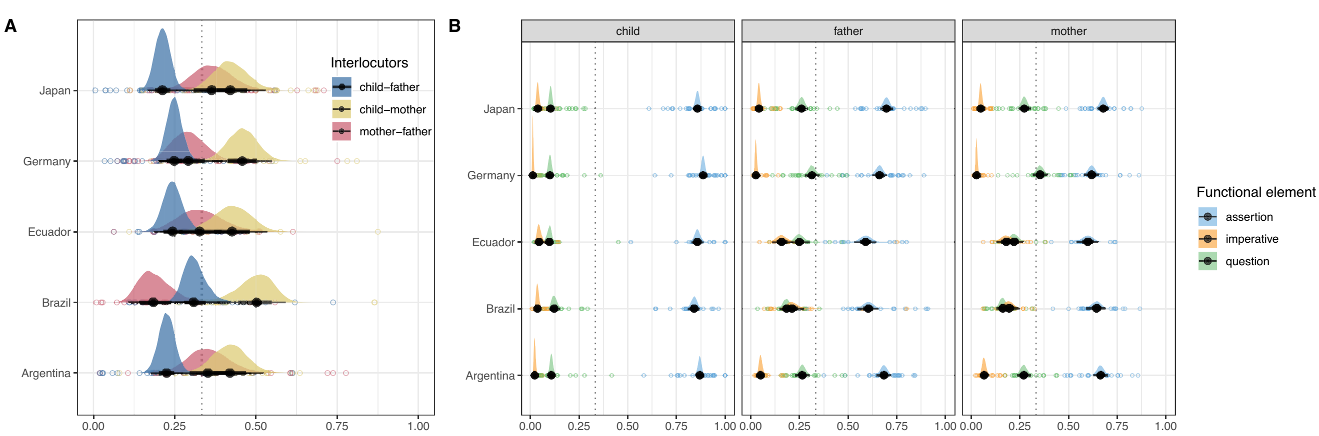
The model comparison clearly favored the interaction model assuming that the use of functional elements varied across speakers within cultural setting, WAIC = 23,591.46, *SE* = 180.20, weight > 0.99; second best model (without interaction): WAIC = 23,689.02, *SE* = 181.03, weight < 0.01; dWAIC = −48.78, *SE*(dWAIC) = 10.42. The general pattern was that assertions were the most frequent type of functional element, followed by questions and imperatives. This ordering was much more pronounced in children in that they hardly used questions or imperatives. Variation across settings was most notable in that both mothers and fathers from Brazil and Ecuador were substantially more likely to use imperatives compared to the other three settings (see Figure 2B).

How Many People Are Involved in a Theme?

Next, we turned to themes as the focus of analysis. As a first step, we asked how many different speakers were involved in a theme. To be involved in a theme, an individual had to produce at least one utterance. Please note that it was possible for themes to have only one speaker. In fact, this was the case for 34% of all utterances. These themes were

Figure 2

Interlocutors and Functional Elements



Note. (A) Proportion of utterances that were exchanged by a pair of interlocutors. Shading shows the interlocutors involved in the utterance regardless of direction (i.e., identity of speaker and listener). (B) Proportion of Utterances That belonged to a certain class of functional element. Facets show different speakers, shading denotes the functional element. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family. CrI = credible intervals. See the online article for the color version of this figure.

mostly single utterances that occurred when someone made an unrelated comment or asked a question but did not receive an answer. We counted the number of speakers involved in each theme (1, 2, or 3) and modeled the resulting distribution using a binomial model. Note that this approach does not take into account the length of each theme. We compared a null model including only an overall intercept `(no_speakers ~ 1)` to a model including cultural setting, `no_speakers ~ 1 + (1|setting)`.

The model comparison favored the model including cultural setting (WAIC = 6,544.58, *SE* = 41.13, weight = 0.95) over the null model (WAIC = 6,550.36, *SE* = 40.88, weight = 0.05; dWAIC = −2.89, *SE*(dWAIC) = 2.77). Figure 3A shows that the number of speakers involved in a theme was relatively similar across cultural settings, with Brazil being the notable exception in having, on average, more speakers per theme.

Who Initiates Themes?

In the following analysis, we asked whether there are differences among speakers and cultural settings in who initiated a theme. For each theme, we only selected the first utterance and used a categorical model to predict the probability with which each individual was the speaker of that utterance and thus the initiator of the theme. Once again, we compared a null model including only an overall intercept `(initiator ~ 1)` to a model including cultural setting `(initiator ~ 1 + (1|setting))`.

The model comparison favored the model including cultural setting (WAIC = 6,566.90, *SE* = 26.07, weight = 0.73) over the null model (WAIC = 6,568.84, *SE* = 25.58, weight = 0.27). However, the difference between models was rather small, suggesting that there were no pronounced differences between cultural settings, dWAIC = −0.97, *SE*(dWAIC) = 1.61. Overall, there were no huge differences between the three individuals in terms of the probability of being the initiator of a theme (range: 0.26–0.41). Compared to an equal split, mothers were slightly more likely to initiate themes and fathers less likely. This relative pattern held for all cultural settings, except Brazil, where the child was the most likely initiator of a theme (see Figure 3B).

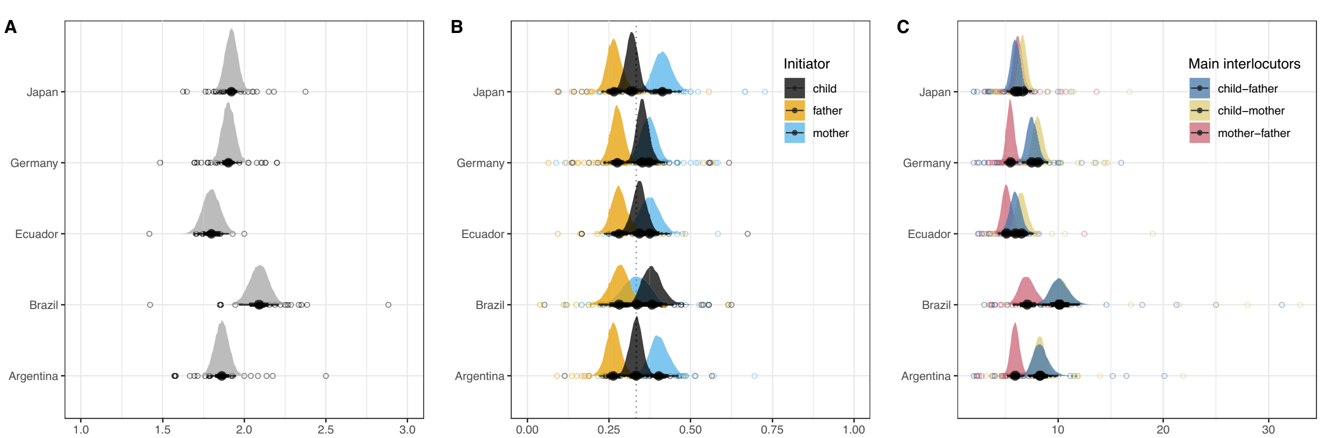
How Long Do Themes Last?

We finished the analysis of themes by asking about variations in how long themes lasted (i.e., how many utterances there were in a theme). For each theme, we noted its length (i.e., the maximum utterance) and the main interlocutors. For that, we counted how many utterances were exchanged between all possible pairs in each theme and classified each theme as being mainly a conversation between those interlocutors who exchanged the most utterances. As a consequence, we excluded all themes that only had a single utterance and only involved a single speaker. The dependent variable (length of the theme) was heavily right-skewed and close to zero and we, therefore, used a log-normal distribution to model it.

The null model only included an intercept and a random effect of family, `theme_length ~ 1 + (1|family)`. There were three alternative models: the first included interlocutors as a fixed effect, `theme_length ~ interlocutors + (1|family)`, the second model added to this a random effect for setting, `theme_length ~ interlocutors + (1|setting) + (1|family)`, and the third model included an additional random slope for interlocutors within setting to model the interaction between interlocutors and setting, `theme_length ~ interlocutors + (interlocutors|setting) + (1|family)`.

The model comparison favored the interaction model assuming that the difference in length of themes for each pair of interlocutors varied across cultural settings—WAIC = 11,657.48, *SE* = 106.30, weight > 0.99; second best model (without interaction): WAIC = 11,671.33, *SE* = 106.59, weight < 0.01; dWAIC = −6.92, *SE*(dWAIC) = 4.11. The average predicted length of a theme across interlocutors and settings was 5.71 utterances, 95% CrI [3.95–8.35]. Figure 3C indicates a variable pattern across cultural settings. In Japan, themes were approximately equally long for all pairs of interlocutors. In the other

Figure 3
Themes



Note. (A) Average number of people involved in a theme. (B) Proportion of themes as a function of who initiated them. Shading shows the initiator. (C) Number of utterances per theme depending on the interlocutors involved. Shading shows the interlocutors who exchanged the most utterances within a given theme. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family. CrI = credible intervals. See the online article for the color version of this figure.

settings, conversations between the mother and father were shorter compared to conversations between one of the parents and the child. This pattern was less pronounced in Ecuador compared to Germany, Brazil, and Argentina. Overall, themes lasted slightly longer in Brazil compared to the other settings.

Family Level Clustering

In this final analysis, we took a more holistic look at the data and tried to identify patterns across the communicative dimensions analyzed above. That is, we asked if there were clusters within our sample that represented different communicative profiles. This allowed us to see (a) if families clustered based on cultural settings and (b) how the different cultural settings clustered with each other. To construct the data set for this analysis, we computed the following dimensions for each family: the amount of `Non-talk`, the proportion of utterances coming from each individual (`Father speaker`, `Mother speaker`, and `Child speaker`), the proportion of `Questions`, `Assertions`, and `Imperatives`, the number of `Gestures`, the number of Themes, the average number of `Utterances` per theme, and the average number of `Speakers per theme`. Please note that more granular dimensions (e.g., gestures or functional elements separate for each individual) would have been possible. However, because this would have meant that each dimension would have had to be estimated based on less data (resulting in a more noisy estimate), we decided to use a more coarse approach.

We performed *k*-means clustering on the data using the function kmeans from the `stats` package which is a native component of R. This analysis partitions the data into *k* clusters so that the sum of squares from points to the assigned cluster centers—in the multidimensional space that is defined by the different dimensions—is minimized. We used the default *Hartigan-Wong* algorithm to find these cluster centers (Hartigan & Wong, 1979). To determine the number of clusters, we used the silhouette and elbow methods via the function `fviz_nbclust` from the `factoextra` package (Kassambara & Mundt, 2020). Both suggested two clusters as the optimal solution.

Figure 4A visualizes the clustering of families based on this analysis. The first cluster (blue), included mainly families from Argentina, Germany, and Japan. Within the cluster, there was no further clustering of families by cultural setting. The second cluster (gold), mainly comprised families from Ecuador and Brazil. Within that cluster, families further tended to cluster by cultural setting, with families from Brazil being more similar to each other compared to families from Ecuador.

In comparison to the first cluster, the second cluster (mainly Ecuador and Brazil) was characterized by overall less talk, a higher proportion of child- compared to parental-talk, and fewer gestures. Furthermore, there were fewer themes, but themes had more speakers and lasted longer. Finally, there was a higher proportion of imperatives and thus fewer assertions and questions (see Figure 4B).

Figure 4C shows the correlations between the different dimensions across clusters. Besides some expected patterns (e.g., negative correlation between the proportion of talk from the different individuals) there were some notable associations: more nontalk was associated with a higher proportion of imperatives, themes had more utterances the more speakers were involved, and more of questions was associated with more themes.

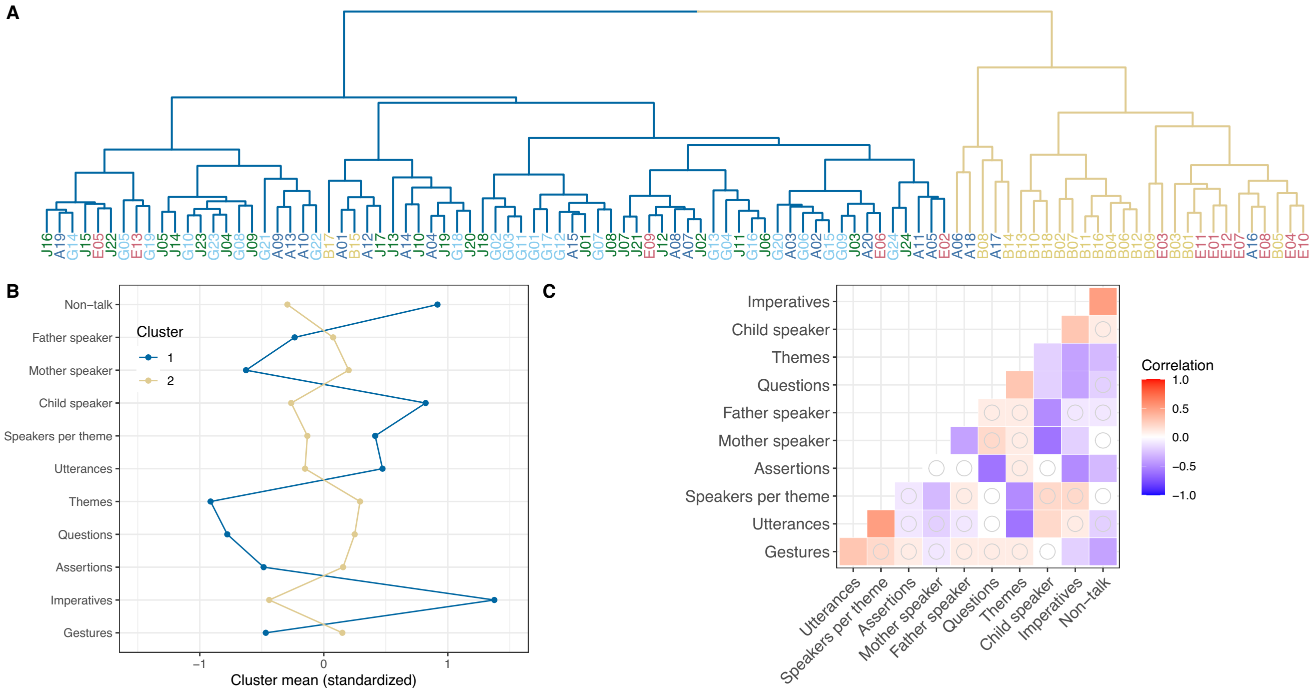
Discussion

We investigated parent–child communicative interactions during mealtimes in five cultural settings. Each family comprised a father, mother, and one child and we analyzed 10 min of video recordings. We found that families from Ecuador and Brazil spent less time talking and used fewer gestures than those from Argentina and Germany, with Japan falling in the middle. Across settings, there was a common pattern in how talk was distributed across family members: mothers talked the most, and children were addressed most frequently. Assertions were the most common type of functional element for all speakers in all settings, followed by questions and imperatives. However, mothers and fathers from Brazil and Ecuador were more likely to use imperatives. The number of themes—parts of coherent utterances—tended to be longer and involved more people in Brazil compared to the other settings. When investigating how families clustered based on their communicative interaction patterns, we found what can be described as an urban-rural split, with families from urban settings (Argentina, Germany, Japan) being more similar to each other compared to families from rural settings (Brazil, Ecuador). These systematic, quantitative comparisons provide an important step toward understanding the similarities and differences in communicative contexts in which children learn language.

Our findings echo how Barrett (2020; see also Kärtner, Schuhmacher, & Giner Torréns, 2020) summarized much of cross-cultural research in the last two decades: variation on a theme. For every aspect of communicative interaction, we investigated, there was a dominant pattern that described behavior in most of the cultural settings but which was often modified in one or two settings. Modification meant that the predicted means for some of the settings were shifted while the distributions of families were largely overlapping. For example, on average, the number of people involved in a theme was around 1.8, with the highest predicted average for Brazil (~2.1) and the lowest for Ecuador (~1.6), yet, the minimum family average in Brazil was 1.40 and the maximum for Ecuador was 2. Similarly, mothers talked the most in all settings but the difference compared to fathers and children was less pronounced in Ecuador and Brazil. Thus, we may tentatively conclude that these overlaps in communicative patterns allow children to use similar learning strategies across settings—in particular, those strategies that leverage the structure of the communicative context (Casillas et al., 2020; Rogoff et al., 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016).

The overall pattern—or theme—can be summarized as being child-centered. Across cultural settings, most talk was directed toward the child. This lends support to theories highlighting the role of direct input in language learning (Braginsky et al., 2019; Goodman et al., 2008; Kachergis et al., 2022; Swingley & Humphrey, 2018). Despite absolute differences in how much input children received, across settings parents directed the largest proportion of talk at the child. Meals are structured by cultural norms that the child has yet to learn, resulting in more direct instruction and—as a by-product—more child-directed linguistic input (Blum-Kulka, 2012; Fjellström, 2004; Köster et al., 2022; Ochs & Shohet, 2006). Furthermore, themes had more conversational turns (i.e., number of utterances) when the child was involved. This finding corresponds well with the idea that children's language learning benefits from coherent and structured interactions (Casillas et al., 2020; Rogoff et al., 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). More frequent conversational turns could originate from adults gradually

Figure 4
Family-Level Clustering



Note. (A) Dendrogram visualizing the similarity between families based on a cluster analysis assuming two clusters. Line shadings show the two clusters, shading of the letters for family corresponds to the different cultural settings. The first letter of the family name denotes the cultural setting (e.g., J = Japan). (B) Mean values for the two clusters for each (standardized) dimension on which the cluster analysis was based. (C) Pearson correlations between the different dimensions entering the cluster analysis. The shading of cells shows the size and direction of the correlation coefficient. Cells without circles show correlations with p values < .05. See the online article for the color version of this figure.

adjusting and elaborating their utterances to the child's response (or lack thereof), resulting in a form of linguistic scaffolding (Bruner, 1983; Vygotsky, 1978). Taken together, the child-centered way of communication might be the consequence of how the interactions in which talk occurs are structured.

Mothers seemed to be the driving force behind this child-centered communicative pattern: they spoke the most, initiated most themes, and most of the themes they were involved in also included the child. This aligns with the former analyses of these videos showing that mothers teach more compared to fathers (Köster et al., 2022) and a recent study by Broesch et al. (2021) who described mothers as the primary interaction partners for young children across five cultural settings. Fathers spoke less and were less likely to be involved in a conversation with the child. As mentioned above, this overall pattern was modified in some of the cultural settings and in the following, we will take a closer look at this variation.

The cluster analyses showed that families' communicative interaction patterns covaried with the degree of urbanization. Families from Brazil and Ecuador were more similar to each other than they were to families from Argentina, Germany, and Japan. Interestingly, within the rural cluster, there seemed to be a further grouping by setting. This was not the case within the urban cluster: even though they lived in very different geographical regions and spoke very different languages. That is, families from Argentina, Germany, and Japan were not more similar to families from the same setting than they were to families from the other settings. However, the urban/rural split was by no means complete in that some of the families from Brazil and Ecuador were assigned to the urban cluster, and some families from Argentina were grouped in the rural cluster. A similar difference between urban and rural settings was found when analyzing parental teaching behavior for these samples but with a stronger subclustering of families in the urban cluster (Köster et al., 2022). Taken together, these results show that variation in communicative interactions did not—at least not primarily—originate from the languages that were spoken but might have been because of norms, values, and beliefs prevalent in the respective cultural settings.

Several theoretical frameworks have focused on different parental socialization goals in urban and rural settings. For example, Keller (2007) described that parents in urban settings prioritize children's independence, while parents in rural settings prioritize interdependence. In line with these proposals, we found that parents in urban samples used more questions, and parents in rural samples used more imperatives, likely reflecting an emphasis on autonomy and compliance, respectively. Furthermore, themes included more speakers in rural contexts, which could reflect a stronger orientation toward the group as opposed to the dyad (Rogoff et al., 2003). Not in line with this general interpretation was the finding that children spoke more in the rural context. Children in urban settings have been described as more communicative because they receive more prompts from their caregivers (Keller, 2007). Below we discuss in more detail how specific norms, values, and beliefs may have influenced the communicative interactions.

Families from Brazil and Ecuador had longer periods of nontalk and produced fewer gestures compared to families from Argentina and Germany. Japanese families fell somewhere in between. This mirrors the results by Cristia (2022) who synthesized 29 studies on naturalistic language input and found that infants growing up in rural settings heard less child-directed speech compared to children growing up in urban settings. It is also in line with the cultural norms that have been described for some of the settings. For the Kichwa community in Ecuador, Sánchez-Parga (2010) reports a norm that meals are supposed to be taken in silence. In Japan, meals are also supposed to be silent under some circumstances (Imada & Furumitsu, 2020). In Germany and Argentina, family meals are seen as a privileged occasion for communication (Aguirre, 2016; Danesi, 2018). In our sample, such norms seemed to have influenced mothers' communication the most: there was less talk by mothers in Ecuador compared to the other settings (except Brazil), while the amount of talk by fathers and children was relatively similar. However, given that all family members talked in all settings, it is worth pointing out that such norms—at least in the present study—mainly had an attenuating effect.

Children communicated in very similar ways across settings: they mostly made assertions and rarely asked questions or used imperatives. Parents' communication in the different settings was also very similar in that they mostly made assertions, asked relatively few questions and hardly used any imperatives. Notably, the rate of imperatives was substantially higher in the rural settings in Brazil and Ecuador. For rural Brazil, Köster et al. (2016) reported that mothers assigned tasks to their children in a more assertive way compared to mothers from urban Germany (see also Keller et al., 2004 for similar findings from rural Costa Rica). Furthermore, when Köster et al. (2022) coded teaching behavior in the same samples, they found that a higher rate of parents in Brazil and Ecuador prompted their children to do something. Finally, in a study on norm enforcement, children living in rural settings themselves used more imperatives than norm-protest when reacting to a peer's perceived norm violation (Kanngiesser et al., 2022). Thus, the higher rate of imperatives might reflect cultural norms and beliefs about how children should behave and how they learn (Keller, 2007).

Limitations

We see the mealtime setting in which we investigated communicative interactions among family members as a strength of the current study, but acknowledge that it comes with important limitations. The constellation of mother, father, and one child is probably more representative of the urban contexts of Argentina, Germany, and Japan than the rural settings. Thus, it would be interesting to see if and how our observed patterns are changed when more people (especially more children and extended family members) take part in the meal. Based on our current findings, we would anticipate similar rates of change across cultural settings. For example, we would expect that the presence of a second child would lower the rate of talk addressed to the other child in a similar way in all cultural settings. Of course, this prediction—as well as all our results—can only generalize to cultural settings in which the interaction format of joint mealtimes exists.

Furthermore, our sample was a convenience sample in that we relied on established contacts and collaborations to recruit families in different settings. As such, the grouping into rural and urban contexts is confounded with the normative belief systems of particular regions. Thus, we do not think that living in a rural setting per se affects communicative interactions in a systematic way, but the specific cultural norms and practices associated with rural subsistence in these settings produced the patterns we observed. Future work should combine our quantitative approach with a qualitative assessment of the local norms surrounding communication and mealtime to better understand the link between norms, values and beliefs, and communicative behavior.

Finally, we did not obtain a measure of children's language abilities. As such, we can only speculate to what extent the different interaction patterns directly affected children's language learning. Obtaining such measures would be a valuable extension of our work.

Conclusions

Our findings offer important insights into the variable and constant aspects of children's language learning environments across diverse cultural settings. For all aspects of communication, we investigated in the current study, a common pattern emerged across cultural settings, suggesting that children can rely on similar information sources and learning processes. This common pattern was modified in some of the settings in a way that might reflect particular local norms, values, beliefs, and ecologies. This exemplifies the importance of quantitative cross-cultural research for theory building in language acquisition.

References

Aguirre, P. (2016). La olla, la fuente y el plato. Distintas maneras de compartir en Argentina [The pot, the platter and the dish: Various ways of sharing in Argentina]. *Studium*, *22*, 189–208. https://doi.org/10.26754/ojs_studium/stud.2016223030

Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language*, *29*(3), 677–686. https://doi.org/10.1017/S0305000902005214

Altınkamış, N. F., Kern, S., & Sofu, H. (2014). When context matters more than language: Verb or noun in French and turkish caregiver speech. *First Language*, *34*(6), 537–550. https://doi.org/10.1177/0142723714560179

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, *41*(5), 430–444. https://doi.org/10.1016/j.evolhumbehav.2020.06.006

Bang, J. Y., Bohn, M., Ramirez, J., Marchman, V. A., & Fernald, A. (2023). Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of later vocabulary than their use of referential gestures. *Developmental Science*, *26*, Article e13354. https://doi.org/10.1111/desc.13354

Bang, J. Y., Mora, A., Munévar, M., Fernald, A., & Marchman, V. A. (2022). *Time to talk: Multiple sources of variability in caregiver verbal engagement during everyday activities in English- and Spanish-speaking families in the US*. PsyArXiv.

Barbaro, K., & Fausey, C. M. (2022). Ten lessons about infants' everyday experiences. *Current Directions in Psychological Science*, *31*(1), 28–33. https://doi.org/10.1177/09637214211059536

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*, *24*(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

Beals, D. E. (1993). Explanatory talk in low-income families' mealtime conversations. *Applied Psycholinguistics*, *14*(4), 489–513. https://doi.org/10.1017/S0142716400010717

Beals, D. E. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language*, *24*(3), 673–694. https://doi.org/10.1017/S0305000997003267

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do north American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), Article e12724. https://doi.org/10.1111/desc.12724

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., & Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *British Journal of Educational Psychology*, *76*(3), 501–515. https://doi.org/10.1348/000709905X49719

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Kleutsch, L., Kramer, K. L., Ross, E., Vongsachang, H., Wrangham, R., Warneken, F., & Bowie, A. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. https://doi.org/10.1038/nature15703

Blum-Kulka, S. (2012). *Dinner talk: Cultural patterns of sociability and socialization in family discourse*. Routledge.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*(1), 223–249. https://doi.org/10.1146/annurev-devpsych-121318-085037

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, *12*(2), 104–108. https://doi.org/10.1111/cdep.12269

Bohn, M., Le, K. N., Peloquin, B., Köymen, B., & Frank, M. C. (2021). Children's interpretation of ambiguous pronouns based on prior discourse. *Developmental Science*, *24*(3), Article e13049. https://doi.org/10.1111/desc.13049

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour*, *5*(8), 1046–1054. https://doi.org/10.1038/s41562-021-01145-1

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Broesch, T., Carolan, P. L., Cebioğlu, S., Rueden, C., Boyette, A., Moya, C., Hewlett, B., & Kline, M. A. (2021). Opportunities for interaction: Natural observations of children's social behavior in five societies. *Human Nature*, *32*(1), 208–238. https://doi.org/10.1007/s12110-021-09393-w

Broesch, T., Lew-Levy, S., Kärtner, J., Kanngiesser, P., & Kline, M. (2022). A roadmap to doing culturally grounded developmental science. *Review of Philosophy and Psychology*, *14*, 587–609. https://doi.org/10.1007/s13164-022-00636-y

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In N. J. Enfield, P. Kockelman, & J. Sidnell (Eds.), *Cambridge handbook of linguistic anthropology* (pp. 187–226). Cambridge University Press.

Bruner, J. (1983). *Child's talk: Learning to use language*. Norton.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., & Migdalek, M. (2020). *A cross-cultural examination of young children's everyday language experiences*. https://doi.org/10.31234/osf.io/723pr

Bürkner, P.-C. (2017). Brms: An r package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Casillas, M. (2022). Learning language in vivo. *Child Development Perspectives*, *17*(1), 10–17. https://doi.org/10.1111/cdep.12469

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a tseltal mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a papuan community. *Journal of Child Language*, *48*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, *27*(1), 69–96. https://doi.org/10.1017/S0305000999004018

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Cristia, A. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *26*(1), Article e13265. https://doi.org/10.1111/desc.13265

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2021). A thorough evaluation of the language environment analysis (LENA) system. *Behavior Research Methods*, *53*(2), 467–486. https://doi.org/10.3758/s13428-020-01393-5

Danesi, G. (2018). A cross-cultural approach to eating together: Practices of commensality among French, German and Spanish young adults. *Social Science Information*, *57*(1), 99–120. https://doi.org/10.1177/0539018417744680

De León, L. (2011). Language socialization and multiparty participation frameworks. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 81–111). Wiley-Blackwell.

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science*, *23*(1), Article e12843. https://doi.org/10.1111/desc.12843

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development*, *92*(2), 609–625. https://doi.org/10.1111/cdev.13511

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. *Language and Linguistics Compass*, *16*(1), Article e12448. https://doi.org/10.1111/lnc3.12448

Fjellström, C. (2004). Mealtime and meal patterns from a cultural perspective. *Scandinavian Journal of Nutrition*, *48*(4), 161–164. https://doi.org/10.1080/11026480410000986

Gaskins, S. (2006). Cultural perspectives on infant–caregiver interaction. In N. J. Enfield & S. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 279–298). Berg.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83–92. https://doi.org/10.1177/1525740110367826

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

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Hewlett, B. S., Lamb, M. E., Shannon, D., Leyendecker, B., & Schölmerich, A. (1998). Culture and early infancy among central African foragers and farmers. *Developmental Psychology*, *34*(4), 653–661. https://doi.org/10.1037/0012-1649.34.4.653

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C., Smith, A. M., Silk, J. B., & Yilmaz, S. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, *4*(1), 36–44. https://doi.org/10.1038/s41562-019-0734-z

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., Hewlett, B. S., McElreath, R., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National*

Academy of Sciences, *110*(36), 14586–14591. https://doi.org/10.1073/pnas.1221217110

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

Imada, S., & Furumitsu, I. (2020). Traditional and modern eating in Japan. In H. Meiselman (Ed.), *Handbook of eating and drinking*. (pp. 1343–1366) Springer.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

Kachergis, G., Marchman, V. A., & Frank, M. C. (2022). Toward a "standard model" of early language learning. *Current Directions in Psychological Science*, *31*(1), 20–27. https://doi.org/10.1177/09637214211057836

Kanngiesser, P., Schäfer, M., Herrmann, E., Zeidler, H., Haun, D., & Tomasello, M. (2022). Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences*, *119*(1), Article e2112521118. https://doi.org/10.1073/pnas.2112521118

Kärtner, J., Schuhmacher, N., & Giner Torréns, M. (2020). Culture and early social-cognitive development. *Progress in Brain Research*, *254*, 225–246. https://doi.org/10.1016/bs.pbr.2020.06.011

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and visualize the results of multivariate data analyses*. https://CRAN.R-project.org/package=factoextra

Keller, H. (2007). *Cultures of infancy*. Lawrence Erlbaum Associates.

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: Self-recognition and self-regulation in three cultural communities. *Child Development*, *75*(6), 1745–1760. https://doi.org/10.1111/j.1467-8624.2004.00814.x

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Köster, M., Cavalcante, L., Vera Cruz de Carvalho, R., Dôgo Resende, B., & Kärtner, J. (2016). Cultural influences on toddlers' prosocial behavior: How maternal task assignment relates to helping others. *Child Development*, *87*(6), 1727–1738. https://doi.org/10.1111/cdev.12636

Köster, M., Giner Torréns, M., Kärtner, J., Itakura, S., Cavalcante, L., & Kanngiesser, P. (2022). Parental teaching behavior in diverse cultural contexts. *Evolution and Human Behavior*, *43*(5), 432–441. https://doi.org/10.1016/j.evolhumbehav.2022.07.002

Kuchirko, Y. A., Schatz, J. L., Fletcher, K. K., & Tamis-Lemonda, C. S. (2020). Do, say, learn: The functions of mothers' speech to infants. *Journal of Child Language*, *47*(1), 64–84. https://doi.org/10.1017/S0305000919000308

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). *An open-source voice type classifier for child-centered daylong recordings*. arXiv Preprint arXiv:2005.12656.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), Article 20130302. https://doi.org/10.1098/rstb.2013.0302

Lieven, E. (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 University Press.

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2021). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*(6), 1–20. https://doi.org/10.1017/S0305000921000623

Masek, L. R., Ramirez, A. G., McMillan, B. T., Hirsh-Pasek, K., & Golinkoff, R. M. (2021). Beyond counting words: A paradigm shift for the study of language acquisition. *Child Development Perspectives*, *15*(4), 274–280. https://doi.org/10.1111/cdep.12425

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in r and stan*. Chapman; Hall/CRC.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Meylan, S. C., & Bergelson, E. (2022). Learning through processing: Toward an integrated approach to early word learning. *Annual Review of Linguistics*, *8*(1), 77–99. https://doi.org/10.1146/annurev-linguistics-031220-011146

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Ochs, E., & Shohet, M. (2006). The cultural structuring of mealtime socialization. *New Directions for Child and Adolescent Development*, *2006*(111), 35–49. https://doi.org/10.1002/cd.154

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, *54*(1), 175–203. https://doi.org/10.1146/annurev.psych.54.101601.145118

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, *2*(1), 201–223. https://doi.org/10.1146/annurev-devpsych-042220-121816

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Sánchez-Parga, J. (2010). *El oficio de antropólogo. Críticia de la razón (inter) cultural*. Ediciones Abya-Yala.

Shneidman, L., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Shneidman, L., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin*, *142*(1), 1–17. https://doi.org/10.1037/bul0000023

Snow, C. E., & Beals, D. E. (2006). Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development*, *2006*(111), 51–66. https://doi.org/10.1002/cd.155

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PLoS ONE*, *8*(11), Article e80646. https://doi.org/10.1371/journal.pone.0080646

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Development*, *90*(4), 1303–1318. https://doi.org/10.1111/cdev.13072

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child Development*, *89*(4), 1247–1267. https://doi.org/10.1111/cdev.12731

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152. https://doi.org/10.1111/cdev.13089

Tamis-LeMonda, C. S., Kuchirko, Y., & Suh, D. D. (2018). Taking center stage: Infants' active role in language learning. In M. M. Saylor, A. P. Ganea (Eds.), *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms* (pp. 39–53). Springer.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565. https://doi.org/10.1017/S030500099700319X

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. MIT Press.

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, *65*(2), 606–621. https://doi.org/10.2307/1131404

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Wittenburg, P., Brugman, H., Russel, A., Klassman, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research*. In Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation. https://archive.mpi.nl/tla/elan

Received December 1, 2022
Revision received July 12, 2023
Accepted November 10, 2023 ■

nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01145-1

Check for updates

OPEN

How young children integrate information sources to infer the meaning of words

Manuel Bohn[1,2,4] ✉, Michael Henry Tessler[3,4], Megan Merrick[2] and Michael C. Frank[2]

Before formal education begins, children typically acquire a vocabulary of thousands of words. This learning process requires the use of many different information sources in their social environment, including their current state of knowledge and the context in which they hear words used. How is this information integrated? We specify a developmental model according to which children consider information sources in an age-specific way and integrate them via Bayesian inference. This model accurately predicted 2–5-year-old children's word learning across a range of experimental conditions in which they had to integrate three information sources. Model comparison suggests that the central locus of development is an increased sensitivity to individual information sources, rather than changes in integration ability. This work presents a developmental theory of information integration during language learning and illustrates how formal models can be used to make a quantitative test of the predictive and explanatory power of competing theories.

Human communicative abilities are unrivalled in the animal kingdom[1–3]. Language—in whatever modality—is the medium that allows humans to collaborate and coordinate in species-unique ways, making it the bedrock of human culture and society[4]. Thus, to absorb the culture around them and become functioning members of society, children need to learn language[5]. A central problem in language learning is referent identification: to acquire the conventional symbolic relation between a word and an object, a child must determine the intended referent of the word. There is no unique cue to reference, however, that can be used across all situations[6]. Instead, referents can only be identified inferentially by reasoning about the speaker's intentions[7–10]. That is, the child has to infer what the speaker is communicating about on the basis of information sources in the utterance's social context.

From early in development, children use several different mechanisms to harness social-contextual information sources[7,9,11]. Children expect speakers to use new words for unknown objects[12–15], to talk about objects that are relevant[16,17], new in context[18,19] or related to the ongoing conversation[20–22]. These different mechanisms, however, have been mainly described and theorized about in isolation. The implied picture of the learning process is that of a 'bag of tricks': mechanisms that operate (and develop) independently from one another[11]. As such, this view of the learning process does not address the complexity of natural social interaction during which many sources of information are present[6,23]. How do children arbitrate between these sources to accurately infer a speaker's intention?

When information integration is studied directly, the focus is mostly on how children interpret or learn words in light of social-contextual information[24–32]. In one classic study[33], children faced a four-compartment (2 × 2) shelf with a ball, a pen and two glasses in it. The speaker, sitting on the opposite side of the display, saw only three of the four compartments: the ball, the pen and one of the glasses. When the speaker asked for "the glass", children had to integrate the semantics of the utterance with the speaker's visual perspective to correctly infer which of the glasses the speaker was referring to. This study advanced our understanding by documenting that preschoolers use both information sources, a finding confirmed by a variety of other work[26,29,31]. Yet these studies neither specify nor test the process by which children integrate different information sources. When interpreting such findings, work in this tradition refers to social-pragmatic theories of language use and learning[9,10,34–36], all of which assume that information is integrated as part of a social inference process but none of which clearly defines the process. As a consequence, we have no explicit and quantitative theory of how different information sources (and word-learning mechanisms) are integrated.

We present a theory of this integration process. Following social-pragmatic theories of language learning[9,10], our theory is based on the following premises: information sources serve different functional roles but are combined as part of an integrated social inference process[34–37]. Children use all available information to make inferences about the intentions behind a speaker's utterance, which then leads them to correctly identify referents in the world and learn conventional word–object mappings. We formalize the computational steps that underlie this inference process in a cognitive model[38–40]. In contrast to earlier modelling work, we treat word learning as the outcome of a social inference process and not just a cross-situational[41,42] or principle-based learning process[43]. In the remainder of this paper, we rigorously test this theory by asking how well it serves the two purposes of any psychological theory: prediction and explanation[44,45]. First, we use the model to make quantitative predictions about children's behaviour in new situations— predictions we test against new data. This form of model testing has been successfully used with adults[38,46] and here we extend it to children. Next, we quantify how well the model explains the integration process by comparing it to alternative models that make different assumptions about whether information is integrated, how it is integrated and how the integration process develops. Alternative models either assume that children ignore some information sources or—in line with a 'bag of tricks' approach—assume that children compute isolated inferences and then weigh their outcome in a posthoc manner.

[1]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [2]Department of Psychology, Stanford University, Stanford, CA, USA. [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]These authors contributed equally: Manuel Bohn, Michael Henry Tessler. ✉e-mail: manuel_bohn@eva.mpg.de

We focus on three information sources that play a central part in theorizing about language use and learning: (1) expectations that speakers communicate in a cooperative and informative manner[12,16,35], (2) shared common ground about what is being talked about in conversation[36,47,48] and (3) semantic knowledge about previously learned word–object mappings[11,49].

Our rational-integration model arbitrates between information sources via Bayesian inference (Fig. 1f gives model formulae). A listener (L_1) reasons about the referent of a speaker's (S_1) utterance. This reasoning is contextualized by the prior probability ρ of each referent. We treat ρ as a conversational prior which originates from the common ground shared between the listener and the speaker. This interpretation follows from the social nature of our experiments (below). From a modelling perspective, ρ can be (and in fact has been) used to capture non-social aspects of a referent, for example its visual salience[38]. To decide between referents, the listener (L_1) reasons about what a rational speaker (S_1) with informativeness α would say given an intended referent. This speaker is assumed to compute the informativity for each available utterance and then choose the most informative one. The informativity of each utterance is given by imagining which referent a listener, who interprets words according to their literal semantics (what we call a literal listener, L_0), would infer on hearing the utterance. Naturally, this reasoning depends on what kind of semantic knowledge θ_j (for object j) the speaker ascribes to the (literal) listener.

Taken together, this model provides a quantitative theory of information integration during language learning. The three information sources operate on different timescales: speaker informativeness is a momentary expectation about a particular utterance, common ground grows over the course of a conversation and semantic knowledge is learned across development. This interplay of timescales has been hypothesized to be an important component of word meaning inference[42,50] and we link these different time-dependent processes together via their proposed impact on model components. Furthermore, the model presents an explicit and substantive theory of development. It assumes that, while children's sensitivity to the individual information sources increases with age, the way integration proceeds remains constant[7,51]. In the model, this is accomplished by creating age-dependent parameters capturing developmental changes in sensitivity to speaker informativeness (α_i, Fig. 1d), the common ground (ρ_i, Fig. 1c) and object-specific semantic knowledge (θ_{ij}, Fig. 1e).

To test the predictive and explanatory power of our model, we designed a word-learning experiment in which we jointly manipulated the three information sources (Fig. 1). Children interacted on a tablet computer with a series of storybook speakers[52]. This situation is depicted in Fig. 1a(4), in which a speaker (here, a frog) appears with a known object (a duck, right) and an unfamiliar object (the diamond-shaped object, left). The speakers used a new word (for example, "wug") in the context of two potential referents and then the child was asked to identify a new instance of the new word, testing their inference about the speaker's intended referent. To vary the strength of the child's inference, we systematically manipulated the familiarity of the known object (from, for example, the highly familiar "duck" to the relatively unfamiliar "pawn") and whether the familiar or new object was new to the speaker (that is, whether it was part of common ground).

This paradigm allows us to examine the integration of the three information sources described above. First, the child may infer that a cooperative and informative speaker[12,16] would have used the word "duck" to refer to the known object (the duck); the fact that the speaker did not say "duck" then suggests that the speaker is most likely referring to a different object (the unfamiliar object). This inference is often referred to as a 'mutual exclusivity' inference[13,15]. Second, the child may draw on what has already been established in the common ground with the speaker. Listeners expect speakers to communicate about things that are new to the common ground[18,19]. Thus, the inference about the new word referring to the unfamiliar object also depends on which object is new in context (Fig. 1a,b(1)–(3)). Finally, the child may use their previously acquired semantic knowledge; that is, how sure they are that the known object is called "duck". If the known object is something less familiar, such as a chess piece (for example, a pawn), a 3-year-old child may draw a weaker inference, if they draw any inference at all[53–55]. Taken together, the child has the opportunity to integrate their assumptions about (1) cooperative communication, (2) their understanding of the common ground and (3) their existing semantic knowledge. In one condition of the experiment, information sources were aligned (Fig. 1a) while in the other they were in conflict (Fig. 1b).

Results

Predicting information integration across development. We tested the model in its ability to predict 2–5-year-old children's judgments about word meaning. We estimated children's ($n = 148$) developing sensitivities to individual information sources in two separate experiments (Experiments 1 and 2; Fig. 1c-e). In Experiment 1, we estimated children's sensitivity to informativeness jointly with their semantic knowledge. In Experiment 2, we estimated sensitivity to common ground. We then generated parameter-free a priori model predictions (developmental trajectories) representing the model's expectations about how children should behave in new situations in which all three information sources had to be integrated. We generated predictions for 24 experimental conditions: 12 objects of different familiarities (requiring different levels of semantic knowledge), with novelty either conflicting or coinciding (Fig. 1). We compared these predictions to newly collected data from $n = 220$ children from the same age range (Experiment 3). All procedures, sample sizes and analyses were preregistered (Methods).

The results showed a very close alignment between model predictions and the data across the entire age range. That is, the average developmental trajectories predicted by the model resembled the trajectories found in the data (Supplementary Fig. 6). With predictions and data binned by child age (in years), the model explained 79% of the variance in the data (Fig. 2a). These results support the assumption of the model that children integrate all three available information sources.

It is still possible, however, that simpler models might make equally good—or even better—predictions. For example, work on children's use of statistical information during morphology learning showed that children's behaviour was best explained by a model that selectively ignored parts of the input[56]. Thus, we formalized the alternative view that children selectively ignore information sources in the form of three lesioned models (Fig. 2b). These models assume that children follow the heuristic 'ignore x' (with x being one of the information sources) when multiple information sources are presented together.

The no-word-knowledge model uses the same model architecture as the rational-integration model. It uses expectations about speaker informativeness and common ground but omits semantic knowledge that is specific to the familiar objects (that is, uses only general semantic knowledge). The model assumes a listener whose inferences do not vary depending on the particular familiar object but only on the age-specific average semantic knowledge (a marker of gross vocabulary size). The no-common-ground model takes in object-specific semantic knowledge and speaker informativeness but ignores common ground information. Instead of assuming that one object has a higher prior probability to be the referent because it is new in context, the listener thinks that both objects are equally likely to be the referent. As a consequence, the listener does not differentiate between situations in which common ground is aligned or in conflict with the other information sources. Finally, according to the no-speaker-informativeness model, the listener does not

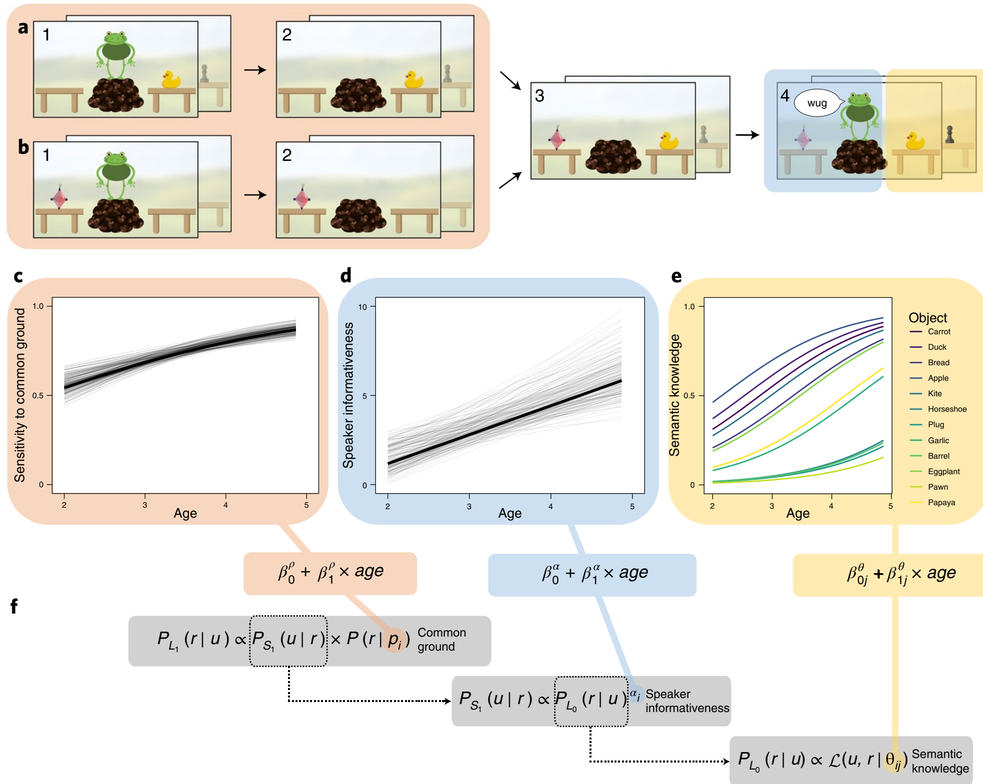


Fig. 1 | Experimental task and model. a,**b**, Screenshots from the experimental task showing the condition of the experiment in which common ground information is congruent (that is, points to the same object) with speaker informativeness (**a**) and also showing the incongruent condition (**b**). The congruent and incongruent conditions are each paired with the 12 known objects, resulting in 24 unique conditions. Steps shown are: the speaker encounters one object and then leaves the scene (1), while the speaker is away (2), a second object appears (3), when returning, the speaker uses a new word to request an object (4). Steps (1) to (3) establish common ground between the speaker and the listener, in that one object is new in context (red). The request in (4) licences an inference based on expectations about how informative speakers are (blue). Listeners' semantic knowledge enters the task because the identity of the known object on one of the tables is varied from well-known objects such as a duck to relatively unfamiliar objects such as a chess pawn (total of 12 objects, yellow). **c**–**e**, Developmental trajectories are shown for sensitivity to common ground (**c**), speaker informativeness (**d**) and semantic knowledge (**e**), estimated on the basis of Experiments 1 and 2 (main text). **f**, The model equation for the rational-integration model, linking information sources to model parameters.

assume that the speaker is communicating in an informative way and hence ignores the utterance. As a consequence, the inference is solely based on common ground expectations.

We found little support for these heuristic models (Fig. 2b). When using Bayesian model comparison via marginal likelihood of the data[57], the data were several orders of magnitude more likely under the rational-integration model compared to any of the lesioned models (rational integration versus no word knowledge: $BF_{10} = 3.9 \times 10^{35}$; rational integration versus no common ground: $BF_{10} = 2.6 \times 10^{47}$; rational integration versus no speaker informativeness: $BF_{10} = 4.8 \times 10^{110}$; Fig. 2). Figure 2c exemplifies the differences between the models: all heuristic models systematically underestimated children's performance in the congruent condition. Thus, even when the information sources were redundant (that is, they all point to the same referent), children's inferences were notably strengthened by each of them. In the incongruent condition, the no-word-knowledge model underestimated performance because it did not differentiate between the different familiar objects. In the case of a highly familiar word such as "duck", it therefore underestimated the effect of the utterance. The no-speaker-informativeness model completely ignored semantic knowledge, which led to even worse predictions. In contrast to the lesioned models that underestimated performance, the no-common-ground model overestimated performance in the incongruent condition because it ignored the dampening effect of common ground favouring the familiar object as the referent. Taken together, we conclude that children considered all available information sources.

Explaining the process of information integration. In the previous section, we established that children integrated all available information sources to infer the meanings of new words. This result, however, does not speak to the process by which information is assumed to be integrated. Thus, in this section, we ask which integration process best explains children's behaviour.

The rational-integration model assumes that all information sources enter into a joint inference process but alternative integration processes are conceivable and might be consistent with the data. For example, the 'bag of tricks'[11] idea mentioned in the introduction could be restated as a modular integration process: children might compute independent inferences on the basis of subsets of the available information and then integrate them in a posthoc manner by weighting them according to some parameter. This view would allow for the possibility that some information sources are considered to be more important than others. In other words, children might be biased towards some information sources. We formalized this alternative view as a biased-integration model. This model assumes that semantic knowledge and expectations about speaker informativeness enter into one inference (mutual exclusivity inference)[12,13,53] while common ground information enters into a second inference. The outcomes of both processes are then weighted according to a

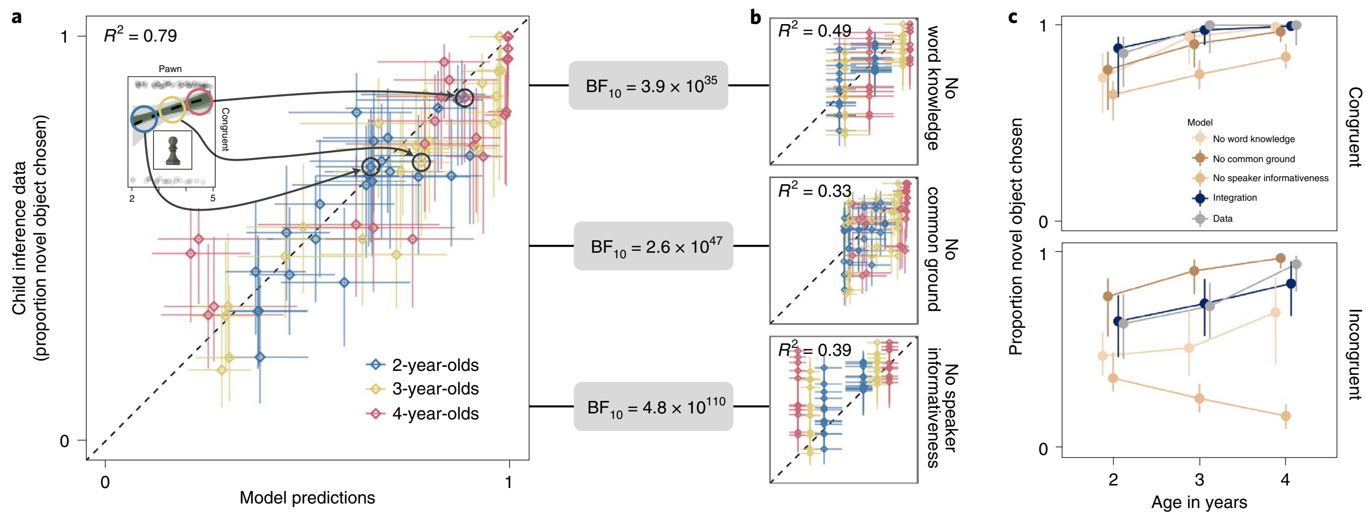


Fig. 2 | Predicting information integration. a,**b**, Correlation between model predictions and child inference data for all 24 conditions and for each age group (binned by year) for the rational-integration model (**a**) and the three lesioned models (**b**). Horizontal and vertical error bars show 95% HDI. Inset in **a** shows an example of model predictions as developmental trajectories (Fig. 3). BF_{10} gives the Bayes factor in favour of the integration model based on the marginal likelihood of the data under each model. **c**, Predictions from all models considered alongside the data (with 95% HDI) for two experimental conditions (familiar word: "duck").

bias parameter ϕ. Like the rational-integration model, this model takes in all available information sources in an age-sensitive way and assumes that they are integrated. The only difference lies in the nature of the integration process: the biased-integration model privileges some information sources over others in an ad-hoc manner.

The parameter ϕ in the biased-integration model is unknown ahead of time and has to be estimated on the basis of the experimental data. That is, through Experiments 1 and 2 alone, we do not learn anything about the relative importance of the information sources. As a consequence—and in contrast to the rational-integration model—the biased-integration model does not allow us to make a-priori predictions about the new data (Experiment 3) in the way we described above. For a fair comparison, we therefore constrained the parameters in the rational-integration model with the data from Experiment 3 as well. As a consequence, both models estimated their parameters using all the data available in a fully Bayesian manner (Supplementary Fig. 4).

The biased-integration model made reasonable posterior predictions and explained 78% of the variance in the data (Fig. 3b). The parameter ϕ—indicating the bias to one of the inferences—was estimated to favour the mutual exclusivity inference (maximum a-posteriori estimate = 0.65; 95% highest density interval (HDI): 0.60–0.71; Fig. 3d). However, the rational-integration model presented a much better fit to the data, both in terms of correlation and the marginal likelihood of the data (Bayes Factor in favour of the rational-integration model: $BF_{10} = 2.1 \times 10^8$; Fig. 3b). When constrained by the data from all experiments, the rational-integration model explained 87% of the variance in the data. Figure 3e exemplifies the difference between the models: the biased-integration model put extra weight on the mutual exclusivity inference and thus failed to capture performance when this inference was weak compared to the common ground inference—such as in the congruent condition for younger children. As a result, a fully integrated—as opposed to a modular and biased—integration process explained the data better.

The rational-integration model assumes that the integration process itself does not change with age[7]. That is, while children's sensitivity to each information source develops, the way the information sources relate to one another remains the same. The biased-integration model can provide the basis for an alternative proposal about developmental change, one in which the integration process itself changes with age. That is, children may be biased towards some information sources and that bias itself may change with age. We formalize such an alternative view as a developmental-bias model which is structurally identical to the biased-integration model but in which the parameter ϕ changes with age. The model assumes that the importance of the different information sources changes with age.

The developmental-bias model also explained a substantial portion of the variance in the data: 78% (Fig. 3c). The estimated developmental trajectory for the bias parameter ϕ suggests that younger children put a stronger emphasis on common ground information, while older children relied more on the mutual exclusivity inference (Fig. 3d). The relative importance of the two inferences seemed to switch at around age 3 yr. Yet again, when we directly compared the competitor models, we found that the data were several orders of magnitude more likely under the rational-integration model (Bayes Factor in favour of the rational-integration model: $BF_{10} = 1.4 \times 10^6$; Fig. 3b). Looking at Fig. 3e, we can see that the developmental-bias model tended to underestimate children's performance because the supportive interplay between the different inferences is constrained. In the biased models, the overall inference could only be as strong as the strongest of the components—in the rational-integration model, the components interacted with one another, allowing for stronger inferences than the individual parts would suggest.

Discussion

The environment in which children learn language is complex. Children have to integrate different information sources, some of which relate to expectations in the moment, others to the dynamics of the unfolding interactions and yet others to their previously acquired knowledge. Our findings show that young children can integrate multiple information sources during language learning—even from relatively early in development. To answer the question of how they do so, we presented a formal cognitive model that assumes that information sources are rationally integrated via Bayesian inference.

Previous work on the study of information integration during language comprehension focused on how adults combine perceptual, semantic or syntactic information[58–62]. Our work extends this work to the development of pragmatics. Our model is based

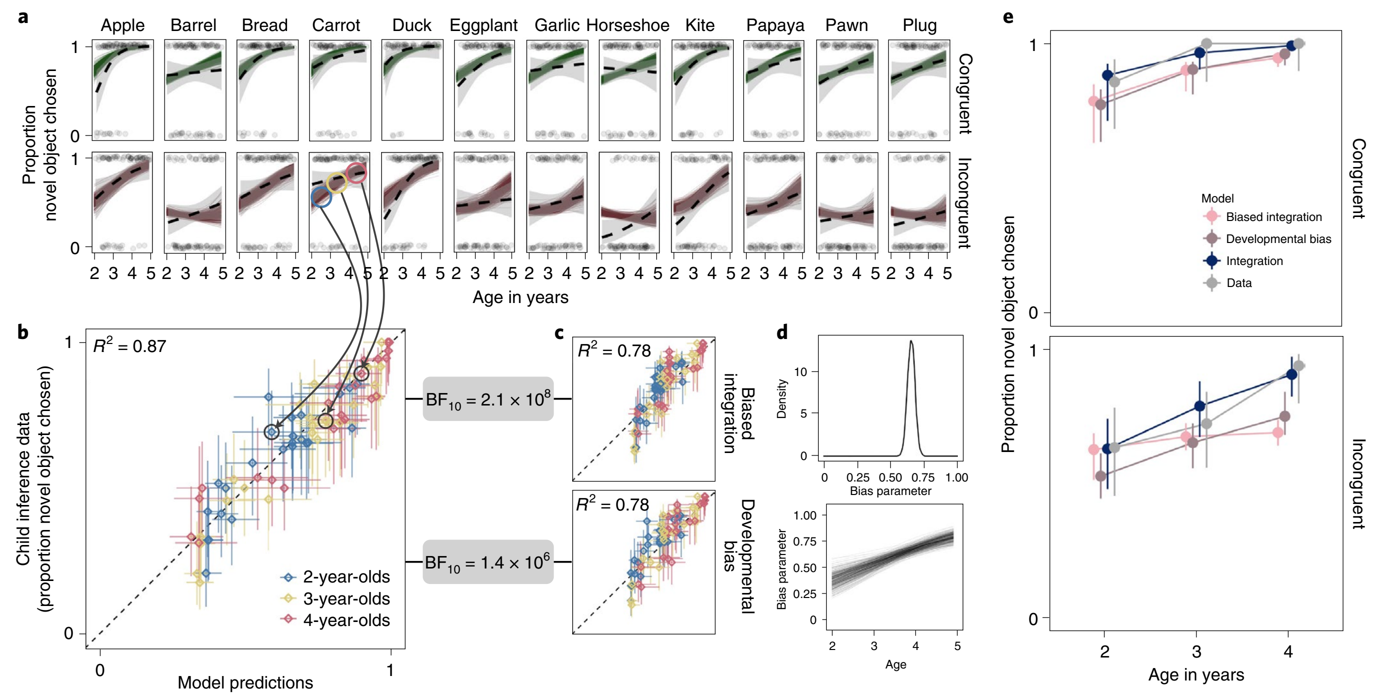


Fig. 3 | Explaining information integration. **a**, Model predictions from the rational-integration model (coloured lines) next to the behavioural data (dotted black lines with 95% CI in grey) for all 24 experimental conditions. Top row shows the congruent condition, while bottom row shows the incongruent condition. Familiar objects are ordered on the basis of their rated age of acquisition (left to right). Light dots represent individual data points. **b**,**c**, Correlations between model predictions binned by age and condition for the integration model (**b**) and the two biased models (**c**). Vertical and horizontal error bars show 95% HDIs. BF_{10} gives the Bayes factor in favour of the rational-integration model on the basis of the marginal likelihood of the data under each model. **d**, Posterior distribution of the bias parameter in the biased-integration model and developmental trajectories for the bias parameter in the developmental-bias model. **e**, Predictions from all models considered alongside the data (with 95% HDI) for two experimental conditions (familiar word: "duck").

on classic social-pragmatic theories on language use and comprehension[10,34–36]. As a consequence, instead of assuming that different information sources feed into separate word-learning mechanisms (the 'bag of tricks' view), we assume that all of these information sources play a functional role in an integrated social inference process. Our model goes beyond previous theoretical and empirical work by describing the computations that underlie this inference process. Furthermore, we presented a substantive theory about how this integration process develops: we assume that children become increasingly sensitive to different information sources but that the way these information sources are integrated remains the same. We used this model to predict and explain children's information integration in a new word-learning paradigm in which they had to integrate (1) their assumptions about informative communication, (2) their understanding of the common ground and (3) their existing semantic knowledge.

The rational-integration model made accurate quantitative predictions across a range of experimental conditions both when information sources were aligned and when they were in conflict. Predictions from the model better explained the data compared to lesioned models which assumed that children ignore one of the information sources, suggesting that children used all available information. We also formalized an alternative, modular, view. According to the biased-integration model, children use all available information sources but compute separate inferences on the basis of a subset of them. Integration happens by weighing the outcomes of these separate inferences by some parameter. Finally, we tested an alternative view on the development of the integration process. According to the developmental-bias model, the importance of the different information sources changes with age. In both cases, the rational-integration model provided a much better fit to the data, suggesting that the integration process remains stable over time. That is, there is developmental continuity and therefore no qualitative difference in how a 2-year-old integrates information compared to a 4-year-old.

The rational-integration model is derived from a more general framework for pragmatic inference, which has been used to explain a wide variety of phenomena in adults' language use and comprehension[38,39,63–67]. Thus, it can be generalized in a natural way to capture word learning in contexts that offer more, fewer or different types of information. For example, non-verbal aspects of the utterance (such as eye-gaze or gestures) can affect children's mutual exclusivity inferences[68–72]. As a first step in this direction, we recently studied how adults and children integrate non-verbal utterances with common ground[51]. Using a structurally similar rational-integration model, we also found a close alignment between model predictions and the data. The flexibility of this modelling framework stems from its conceptualization of human communication as a form of rational social action. As such, it connects to computational and empirical work that tries to explain social reasoning by assuming that humans expect each other to behave in a way that maximizes the benefits and minimizes the cost associated with actions[28,73,74].

Our model and empirical paradigm provide a foundation on which to test deeper questions about language development. First, our experiments should be performed in children from different cultural backgrounds learning different languages[75]. In such studies, we would not expect our results to replicate in a strict sense; that is, we would not expect to see the same developmental trajectories in all cultures and languages. Substantial variation is much more likely. Studies on children's pragmatic inferences in different cultures have documented both similar[76,77] and different[78] developmental trajectories. Nevertheless, our model provides a way to think about how to reconcile cross-cultural variation with a shared cognitive architecture: we predict differences in how sensitive children are to the individual information sources at different ages but similarities in how information is integrated[7]. In computational terms, we assume a universal architecture that specifies the relation between a set of varying parameters. Of course, either confirmation or disconfirmation of this prediction would be informative.

Second, it would be useful to flesh out the cognitive processes that underlie reasoning about common ground. The basic assumption that common ground changes interlocutors' expectations about what are likely referents[79] has been used in earlier modelling work on the role of common ground in reference resolution[62]. Here, we went one step further and measured the strength of these expectations to inform the parameter values in our model. However, in its current form, our model treats common ground as a conversational prior and does not specify how the listener arrives at the expectation that some objects are more likely to be referents because they are new in common ground. That is, computationally, our model does not differentiate between common ground information and other reasons that might make an object contextually more salient. An interesting starting point to overcome this shortcoming would be modelling work on the role of common ground in conversational turn-taking[80].

Finally, our model is a model of referent identification in the moment of the utterance. At the same time, the constructs made use of by our model are shaped by factors that unfold across multiple time points and contexts: common ground is built over the course of a conversation and the lexical knowledge of a child is shaped across a language developmental timescale. Even speaker informativeness could be imagined to vary over time following repeated interactions with a particular speaker. What is more, assessing speaker informativeness is unlikely to be the outcome of a single, easy-to-define process. The expectations about informative communication that we take it to represent are probably the result of the interplay between multiple social and non-social inference processes. The broader point here is that our model makes use of unidimensional representations of high-dimensional, structured processes and examines how these representations are integrated. As such, it is first and foremost a computational description of the inferences and we therefore make no strong claims about the psychological reality of the parameters in it. Connecting our model with other frameworks that focus on the cognitive, temporal and cross-situational aspects of word learning would elucidate further these complex processes[42,50,81].

This work advances our understanding of how children navigate the complexity of their learning environment. Methodologically, it illustrates how computational models can be used to test theories; from a theoretical perspective, it adds to broader frameworks that see the ontogenetic and phylogenetic emergence of language as deeply rooted in social cognition.

Methods

A more detailed description of the experiments and the models can be found in the Supplementary Information. The experimental procedure, sample sizes and analysis for each experiment were preregistered (https://osf.io/7rg9j/registrations; dates of registration: 2 May 2019, 5 April 2019 and 2 March 2019). Experimental procedures, data, model and analysis scripts can be found in an online repository (https://github.com/manuelbohn/spin). Experiments 1 and 2 were designed to estimate children's developing sensitivity to each information source. The results of these experiments determine the parameter values in the model (Fig. 1c-f). Experiment 3 was designed to test how children integrate different information sources.

Participants. Sample sizes for each experiment were chosen to have at least 30 data points per cell (that is, unique combination of condition, familiar object and age group). Across the three experiments, a total of 368 children participated. Experiment 1 involved 90 children, including 30 2-year-olds (range = 2.03–3.00, 15 girls), 30 3-year-olds (range = 3.03–3.97, 22 girls) and 30 4-year-olds (range = 4.03–4.90, 16 girls). Data from ten additional children were not included because they were either exposed to less than 75% of English at home (five children), did not finish at least half of the test trials (two children), the technical equipment failed (two children) or their parents reported an autism spectrum disorder (one child).

In Experiment 2, we tested 58 children, including 18 2-year-olds (range = 2.02–2.93, seven girls), 19 3-year-olds (range = 3.01–3.90, 14 girls) and 21 4-year-olds (range = 4.07–4.93, 14 girls). Data from five additional children were not included because they were either exposed to less than 75% of English at home (three children) or the technical equipment failed (two children).

Finally, Experiment 3 involved 220 children, including 76 2-year-olds (range = 2.04–2.99, seven girls), 72 3-year-olds (range = 3.00–3.98, 14 girls) and 72 4-year-olds (range = 4.00–4.94, 14 girls). Data from 20 additional children were not included because they were either exposed to less than 75% of English at home (15 children), did not finish at least half of the test trials (three children) or the technical equipment failed (two children).

All children were recruited in a children's museum in San José, California, United States. This population is characterized by a diverse ethnic background (predominantly White, Asian or mixed-ethnicity) and high levels of parental education and socioeconomic status. Parents consented to their children's participation and provided demographic information. All experiments were approved by the Stanford Institutional Review Board (protocol no. 19960).

Materials. All experiments were presented as an interactive picture book on a tablet computer. Tablet-based storybooks are commonly used to simulate social interactions in developmental research and interventions[82]. A recent, direct comparison found similar performance with tablet-based and printed storybooks in a word-learning paradigm[52]. Furthermore, our results in Experiment 1 and 2 replicate earlier studies on mutual exclusivity and discourse novelty that used live interactions instead of storybooks[18,19].

Figure 1a,b show screenshots from the actual experiments. The general setup involved an animal standing on a little hill between two tables. For each animal character, we recorded a set of utterances (one native English speaker per animal) that were used to talk to the child and make requests. Each experiment started with two training trials in which the speaker requested known objects (car and ball).

Procedure. Experiment 1 tested the mutual exclusivity inference[13,53]. On one table, there was a familiar object; on the other table, there was an unfamiliar object (a new design drawn for the purpose of the study) (Fig. 1a/b(4) and Supplementary Fig. 1a). The speaker requested an object by saying "Oh cool, there is a (non-word) on the table, how neat, can you give me the (non-word)?". Children responded by touching one of the objects. The location of the unfamiliar object (left or right table) and the animal character were counterbalanced. We coded a response as a correct choice if children chose the unfamiliar object as the referent of the new word. Each child completed 12 trials, each with a different familiar and a different unfamiliar object. We used familiar objects that we expected to vary along the dimension of how likely children were to know the word for it. This set included objects that most 2-year-olds can name (for example, a duck) as well as objects that only very few 5-year-olds can name (for example, a pawn (chess piece)). The selection was based on the age of acquisition ratings from Kuperman and colleagues[83]. While these ratings usually do not capture the absolute age when children acquire these words, they capture the relative order in which words are learned. Supplementary Fig. 2a shows the words and objects used in the experiment. There was a high correlation between the rated age-of-acquisition and the mutual exclusivity effect for the different words (Supplementary Fig. 2c).

Experiment 2 tested children's sensitivity to common ground that is built up over the course of a conversation. In particular, we tested whether children keep track of which object is new to a speaker and which they have encountered previously[18,19]. The general setup was the same as in Experiment 1 (Supplementary Fig. 1b). The speaker was positioned between the tables. There was an unfamiliar object (drawn for the purpose of the study) on one of the tables while the other table was empty. Next, the speaker turned to one of the tables and either commented on the presence ("Aha, look at that.") or the absence ("Hm, nothing there.") of an object. Then the speaker disappeared. While the speaker was away, a second unfamiliar object appeared on the previously empty table. Then the speaker returned and requested an object in the same way as in Experiment 1. The positioning of the unfamiliar object at the beginning of the experiment, the speaker as well as the location the speaker turned to first was counterbalanced. Children completed ten trials, each with a different pair of unfamiliar objects. We coded a response as a correct choice if children chose as the referent of the new word the object that was new to the speaker.

Experiment 3 combined the procedures from Experiments 1 and 2. It followed the same procedure as Experiment 2 but involved the same objects as Experiment 1 (Fig. 1(1)–(4)) and Supplementary Fig. 1c). In the beginning, one table was empty while there was an object (unfamiliar or familiar) on the other one. After commenting on the presence or absence of an object on each table, the speaker disappeared and a second object appeared (familiar or unfamiliar). Next, the speaker reappeared and made the usual request ("Oh cool, there is a (non-word) on the table, how neat, can you give me the (non-word)?"). In the congruent condition, the familiar object was present in the beginning and the unfamiliar object appeared while the speaker was away (Fig. 1a and Supplementary Fig. 1c, left). In this case, both the mutual exclusivity and the common ground inference pointed to the new object as the referent (that is, it was both new to the speaker in the context and it was an object that does not have a label in the lexicon). In the incongruent condition, the unfamiliar object was present in the beginning and the familiar object appeared later. In this case, the two inferences pointed to different objects (Fig. 1b and Supplementary Fig. 1c, right). This resulted in a total of two alignments (congruent versus incongruent) × 12 familiar objects = 24 different conditions. Participants received up to 12 test trials, six in each alignment condition, each with a different familiar and unfamiliar object. Familiar objects were the same as in Experiment 1. The positioning of the objects on the tables, the speaker and the location the speaker first turned to were counterbalanced. Participants could stop the experiment after six trials (three per alignment condition). If a participant stopped after half of the trials, we tested an additional participant from the same age group to reach the preregistered number of data points per age group (2-, 3- and 4-year-olds).

Data analysis. To analyse how the manipulations in each experiment affected children's behaviour, we used generalized linear mixed models. Since the focus of the paper is on how information sources were integrated, we discuss these models in the Supplementary Information and focus here on the cognitive models instead. A detailed, mathematical description of the different cognitive models along with details about estimation procedures and priors can be found in the Supplementary Information. All cognitive models and Bayesian data analytic models were implemented in the probabilistic programming language WebPPL[84]. The corresponding model code can be found in the associated online repository. Information about priors for parameter estimation and Markov chain Monte Carlo settings can also be found in the Supplementary Information and the online repository.

As a first step, we used the data from Experiments 1 and 2 to estimate children's developing sensitivity to each information source. To estimate the parameters for semantic knowledge (θ) and speaker informativeness (α), we adapted the rational-integration model to model a situation in which both objects (new and familiar) have equal prior probability (that is, no common ground information). We used the data from Experiment 1 to then infer the semantic knowledge and speaker informativeness parameters in an age-sensitive manner. Specifically, we inferred the intercepts and slopes for speaker informativeness via a linear regression submodel and semantic knowledge via a logistic regression submodel, the values of which were then combined in the cognitive model to generate model predictions to predict the responses generated in Experiment 1. To estimate the parameters representing sensitivity to common ground (ρ), we used a simple logistic regression to infer which combination of intercept and slope would generate predictions that corresponded to the average proportion of correct responses measured in Experiment 2. For the 'prediction' models, the parameters whose values were inferred by the data from Experiments 1 and 2 were then used to make out-of-sample predictions for Experiment 3. For the 'explanation' models, these parameters were additionally constrained by the data from Experiment 3. A more detailed description of how these parameters were estimated (including a graphical model, Supplementary Fig. 4) can be found in the Supplementary Information.

To generate model predictions, we combined the parameters according to the respective model formula. As mentioned above, common ground information could either be aligned or in conflict with the other information sources. In the congruent condition, the unfamiliar object was also new in context and thus had the prior probability ρ. In the incongruent condition, the new object was the 'old' object and thus had the prior probability of $1-\rho$.

The rational-integration model is a mapping from an utterance u to a referent r, defined as

$$P_{L_1}^{\mathrm{int}}(r \mid u;\{\rho_i, \alpha_i\, \theta_{ij}\}) \propto P_{S_1}(u \mid r;\{\alpha_i, \theta_{ij}\}) \times P(r \mid \rho_i)$$

where i represents the age of the participant and the j the familiar object. The three lesioned models that were used to compare how well the model predicts new data are reduced versions of this model. The no-word-knowledge model uses the same model architecture

$$P_{L_1}^{\mathrm{no\_wk}}(r \mid u;\{\rho_i, \alpha_i\, \theta_i\}) \propto P_{S_1}(u \mid r;\{\alpha_i, \theta_i\}) \times P(r \mid \rho_i)$$

and the only difference lies in the parameter θ, which does not vary as a function of j, the object (that is, θ in this model is analogous to a measure of gross vocabulary development). The object-specific parameters for semantic knowledge are fitted via a hierarchical regression (mixed effects) model. That is, there is an overall developmental trajectory for semantic knowledge (main effect, θ_i) and then there is object-specific variation around this trajectory (random effects, θ_{ij}). Thus, the no-word-knowledge model takes in the overall trajectory for semantic knowledge (θ_i) but ignores object-specific variation. The no-common-ground model ignores common ground information (represented by ρ) and is thus defined as

$$P_{L_1}^{\mathrm{no\_cg}}(r \mid u;\{\alpha_i\, \theta_{ij}\}) \propto P_{S_1}(u \mid r;\{\alpha_i, \theta_{ij}\}).$$

For the no-speaker-informativeness model, the parameter $\alpha = 0$. As a consequence, the likelihood term in the model is 1 and the model therefore reduces to

$$P_{L_1}^{\mathrm{no\_si}}(r \mid u;\{\rho_i\}) \propto P(r \mid \rho_i).$$

As noted above, the explanation models used parameters that were additionally constrained by the data from Experiment 3 but the way these parameters were combined in the rational-integration model was the same as above. The biased-integration model is defined as

$$P_{L_1}^{\mathrm{biased}}(r \mid u;\{\phi, \rho_i, \alpha_i, \theta_{ij}\}) = \phi \times P_{\mathrm{ME}}(r \mid u;\{\alpha_i, \theta_{ij}\}) + (1-\phi) \times P(r \mid \rho_i)$$

with P_{ME} representing a mutual exclusivity inference which takes in speaker informativeness and object-specific semantic knowledge. This inference is then weighted by the parameter ϕ and added to the respective prior probability, which is weighted by $1-\phi$. Thus, ϕ represents the bias in favour of the mutual exclusivity inference. In the developmental-bias model the parameter ϕ is made to change with age (ϕ) and the model is thus defined as

$$P_{L_1}^{\mathrm{dev\_bias}}(r \mid u;\{\phi_i, \rho_i, \alpha_i, \theta_{ij}\}) = \phi_i \times P_{\mathrm{ME}}(r \mid u;\{\alpha_i, \theta_{ij}\}) + (1-\phi_i).$$

We compared models in two ways. First, we used Pearson correlations between model predictions and the data. For this analysis, we binned the model predictions and the data by age in years and by the type of familiar object (Figs. 2 and 3 and Supplementary Figs. 7 and 10). Second, we compared models on the basis of the marginal likelihood of the data under each model—the likelihood of the data averaging over ('marginalizing over') the prior distribution on parameters; the pairwise ratio of marginal likelihoods for two models is known as the Bayes Factor. It is interpreted as how many times more likely the data are under one model compared to the other. Bayes Factors quantify the quality of predictions of a model, averaging over the possible values of the parameters of the models (weighted by the prior probabilities of those parameter values); by averaging over the prior distribution on parameters, Bayes Factors implicitly take into account model complexity because models with more parameters will tend to have a broader prior distribution over parameters, which in effect, can water down the potential gains in predictive accuracy that a model with more parameters can achieve[57]. For this analysis, we treated age continuously.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data files, along with all experimental stimuli, model and analysis scripts can be found at: https://github.com/manuelbohn/spin.

Received: 23 October 2020; Accepted: 27 May 2021;
Published online: 01 July 2021

References

1. Taylor, C. *The Language Animal* (Harvard Univ. Press, 2016).
2. Tomasello, M. *Origins of Human Communication* (MIT Press, 2008).
3. Fitch, W. T. *The Evolution of Language* (Cambridge Univ. Press, 2010).
4. Smith, J. M. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, 1997).
5. Tomasello, M. *A Natural History of Human Thinking* (Harvard Univ. Press, 2018).
6. Frank, M. C., Tenenbaum, J. B. & Fernald, A. Social and discourse contributions to the determination of reference in cross-situational word learning. *Lang. Learn. Dev.* **9**, 1–24 (2013).
7. Bohn, M. & Frank, M. C. The pervasive role of pragmatics in early language. *Annu. Rev. Dev. Psychol.* **1**, 223–249 (2019).
8. Bruner, J. *Child's Talk: Learning to Use Language* (Norton, 1983).
9. Clark, E. V. *First Language Acquisition* (Cambridge Univ. Press, 2009).
10. Tomasello, M. *Constructing a Language* (Harvard Univ. Press, 2009).
11. Bloom, P. *How Children Learn the Meanings of Words* (MIT Press, 2002).
12. Clark, E. V. In *Mechanisms of Language Acquisition* (ed. MacWhinney, B.) 1–33 (Lawrence Erlbaum Associates, 1987).
13. Markman, E. M. & Wachtel, G. F. Children's use of mutual exclusivity to constrain the meanings of words. *Cogn. Psychol.* **20**, 121–157 (1988).
14. Carey, S. & Bartlett, E. Acquiring a single new word. *Proc. Stanf. Child Lang. Conf.* **15**, 17–29 (1978).
15. Halberda, J. The development of a word-learning strategy. *Cognition* **87**, B23–B34 (2003).
16. Frank, M. C. & Goodman, N. D. Inferring word meanings by assuming that speakers are informative. *Cogn. Psychol.* **75**, 80–96 (2014).
17. Schulze, C., Grassmann, S. & Tomasello, M. 3-year-old children make relevance inferences in indirect verbal communication. *Child Dev.* **84**, 2079–2093 (2013).

18. Akhtar, N., Carpenter, M. & Tomasello, M. The role of discourse novelty in early word learning. *Child Dev.* **67**, 635–645 (1996).
19. Diesendruck, G., Markson, L., Akhtar, N. & Reudor, A. Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Dev. Sci.* **7**, 33–41 (2004).
20. Bohn, M., Le, K., Peloquin, B., Koymen, B. & Frank, M. C. Children's interpretation of ambiguous pronouns based on prior discourse. *Dev. Sci.* **24**, e13049 (2021).
21. Sullivan, J. & Barner, D. Discourse bootstrapping: preschoolers use linguistic discourse to learn new words. *Dev. Sci.* **19**, 63–75 (2016).
22. Horowitz, A. C. & Frank, M. C. Young children's developing sensitivity to discourse continuity as a cue for inferring reference. *J. Exp. Child Psychol.* **129**, 84–97 (2015).
23. Bergelson, E. & Aslin, R. N. Nature and origins of the lexicon in 6-mo-olds. *Proc. Natl Acad. Sci. USA* **114**, 12916–12921 (2017).
24. Hollich, G. J., Hirsh-Pasek, K. & Golinkoff, R. M. Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monogr. Soc. Res. Child Dev.* **65**, 17–29 (2000).
25. Ganea, P. A. & Saylor, M. M. Infants' use of shared linguistic information to clarify ambiguous requests. *Child Dev.* **78**, 493–502 (2007).
26. Graham, S. A., San Juan, V. & Khu, M. Words are not enough: how preschoolers' integration of perspective and emotion informs their referential understanding. *J. Child Lang.* **44**, 500–526 (2017).
27. Grosse, G., Moll, H. & Tomasello, M. 21-month-olds understand the cooperative logic of requests. *J. Pragmat.* **42**, 3377–3383 (2010).
28. Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B. & Schulz, L. E. Social pragmatics: preschoolers rely on commonsense psychology to resolve referential underspecification. *Child Dev.* **91**, 1135–1149 (2020).
29. Khu, M., Chambers, C. G. & Graham, S. A. Preschoolers flexibly shift between speakers' perspectives during real-time language comprehension. *Child Dev.* **91**, e619–e634 (2020).
30. Matthews, D., Lieven, E., Theakston, A. & Tomasello, M. The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Appl. Psycholinguist.* **27**, 403–422 (2006).
31. Nilsen, E. S., Graham, S. A. & Pettigrew, T. Preschoolers' word mapping: the interplay between labelling context and specificity of speaker information. *J. Child Lang.* **36**, 673–684 (2009).
32. Kachel, G., Hardecker, D. J. K. & Bohn, M. Young children's developing ability to integrate gestural and emotional cues. *J. Exp. Child Psychol.* **201**, 104984 (2021).
33. Nadig, A. S. & Sedivy, J. C. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychol. Sci.* **13**, 329–336 (2002).
34. Grice, H. P. *Studies in the Way of Words* (Harvard Univ. Press, 1991).
35. Sperber, D. & Wilson, D. *Relevance: Communication and Cognition* (Blackwell, 2001).
36. Clark, H. H. *Using Language* (Cambridge Univ. Press, 1996).
37. Shafto, P., Goodman, N. D. & Frank, M. C. Learning from others: the consequences of psychological reasoning for human learning. *Perspect. Psychol. Sci.* **7**, 341–351 (2012).
38. Frank, M. C. & Goodman, N. D. Predicting pragmatic reasoning in language games. *Science* **336**, 998 (2012).
39. Goodman, N. D. & Frank, M. C. Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci.* **20**, 818–829 (2016).
40. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* **26**, 1596–1618 (2019).
41. Fazly, A., Alishahi, A. & Stevenson, S. A probabilistic computational model of cross-situational word learning. *Cogn. Sci.* **34**, 1017–1063 (2010).
42. Frank, M. C., Goodman, N. D. & Tenenbaum, J. B. Using speakers' referential intentions to model early cross-situational word learning. *Psychol. Sci.* **20**, 578–585 (2009).
43. Xu, F. & Tenenbaum, J. B. Word learning as Bayesian inference. *Psychol. Rev.* **114**, 245 (2007).
44. Shmueli, G. To explain or to predict? *Stat. Sci.* **25**, 289–310 (2010).
45. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
46. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
47. Bohn, M. & Koymen, B. Common ground and development. *Child Dev. Perspect.* **12**, 104–108 (2018).
48. Clark, E. V. in *The Handbook of Language Emergence* (eds. MacWhinney, B. & O'Grady, W.) 328–353 (John Wiley & Sons, 2015).
49. Fenson, L. et al. Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* **59**, 1–173 (1994).
50. McMurray, B., Horst, J. S. & Samuelson, L. K. Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychol. Rev.* **119**, 831 (2012).
51. Bohn, M., Tessler, M. H., Merrick, M. & Frank, M. C. Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/xma4f (2019).
52. Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L. & Yurovsky, D. Using tablets to collect data from young children. *J. Cogn. Dev.* **17**, 1–17 (2016).
53. Lewis, M. L., Cristiano, V., Lake, B. M., Kwan, T. & Frank, M. C. The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition* **198**, 104191 (2020).
54. Grassmann, S., Schulze, C. & Tomasello, M. Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Front. Psychol.* **6**, 1200 (2015).
55. Ohmer, X., König, P. & Franke, M. Reinforcement of semantic representations in pragmatic agents leads to the emergence of a mutual exclusivity bias. In *Proc. 42nd Annual Meeting of the Cognitive Science Society* (eds Denison., S., Mack, M., Xu, Y. & Armstrong, B. C.) 1779–1785 (Cognitive Science Society, 2020).
56. Gagliardi, A., Feldman, N. H. & Lidz, J. Modeling statistical insensitivity: sources of suboptimal behavior. *Cogn. Sci.* **41**, 188–217 (2017).
57. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge Univ. Press, 2014).
58. Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. Integration of visual and linguistic information in spoken language comprehension. *Science* **268**, 1632–1634 (1995).
59. Kamide, Y., Scheepers, C. & Altmann, G. T. Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *J. Psycholinguist. Res.* **32**, 37–55 (2003).
60. Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. Integration of word meaning and world knowledge in language comprehension. *Science* **304**, 438–441 (2004).
61. Özyürek, A., Willems, R. M., Kita, S. & Hagoort, P. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* **19**, 605–616 (2007).
62. Heller, D., Parisien, C. & Stevenson, S. Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition* **149**, 104–120 (2016).
63. Monroe, W., Hawkins, R. X., Goodman, N. D. & Potts, C. Colors in context: a pragmatic neural model for grounded language understanding. *Trans. Assoc. Comput. Linguist.* **5**, 325–338 (2017).
64. Wang, S., Liang, P. & Manning, C. D. Learning language games through interaction. *Proc. 54th Annual Meeting of the Association for Computational Linguistics* (eds Erk, K. & Smith, N. A.) 2368–2378 (Association for Computational Linguistics, 2016).
65. Tessler, M. H. & Goodman, N. D. The language of generalization. *Psychol. Rev.* **126**, 395–436 (2019).
66. Yoon, E. J., Tessler, M. H., Goodman, N. D. & Frank, M. C. Polite speech emerges from competing social goals. *Open Mind* **4**, 71–87 (2020).
67. Savinelli, K., Scontras, G. & Pearl, L. Modeling scope ambiguity resolution as pragmatic inference: formalizing differences in child and adult behavior. *Proc. 39th Annual Meeting of the Cognitive Science Society* (eds Gunzelmann, G., Howes, A., Tenbrink, T. & Davelaar, E. J.) 3064–3069 (Cognitive Science Society, 2017).
68. Paulus, M. & Fikkert, P. Conflicting social cues: 14- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *J. Cogn. Dev.* **15**, 43–59 (2014).
69. Graham, S. A., Nilsen, E. S., Collins, S. & Olineck, K. The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *Br. J. Dev. Psychol.* **28**, 449–465 (2010).
70. Gangopadhyay, I. & Kaushanskaya, M. The role of speaker eye gaze and mutual exclusivity in novel word learning by monolingual and bilingual children. *J. Exp. Child Psychol.* **197**, 104878 (2020).
71. Jaswal, V. K. & Hansen, M. B. Learning words: children disregard some pragmatic information that conflicts with mutual exclusivity. *Dev. Sci.* **9**, 158–165 (2006).
72. Grassmann, S. & Tomasello, M. Young children follow pointing over words in interpreting acts of reference. *Dev. Sci.* **13**, 252–263 (2010).
73. Bridgers, S., Jara-Ettinger, J. & Gweon, H. Young children consider the expected utility of others' learning to decide what to teach. *Nat. Hum. Behav.* **4**, 144–152 (2020).
74. Jara-Ettinger, J., Gweon, H., Schulz, L. E. & Tenenbaum, J. B. The naive utility calculus: computational principles underlying commonsense psychology. *Trends Cogn. Sci.* **20**, 589–604 (2016).
75. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
76. Su, Y. E. & Su, L.-Y. Interpretation of logical words in Mandarin-speaking children with autism spectrum disorders: uncovering knowledge of semantics and pragmatics. *J. Autism Dev. Disord.* **45**, 1938–1950 (2015).

77. Zhao, S., Ren, J., Frank, M. C. & Zhou, P. The development of quantity implicatures in Mandarin-speaking children. *Lang. Learn. Dev.* https://doi.org/10.1080/15475441.2021.1886935 (2021).
78. Fortier, M., Kellier, D., Flecha, M. F. & Frank, M. C. Ad-hoc pragmatic implicatures among Shipibo-Konibo children in the Peruvian Amazon. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/x7ad9 (2018).
79. Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* **49**, 43–61 (2003).
80. Anderson, C. J. Tell me everything you know: a conversation update system for the rational speech acts framework. *Proc. Soc. Comput. Linguist.* **4**, 244–253 (2021).
81. Yurovsky, D. & Frank, M. C. An integrative account of constraints on cross-situational learning. *Cognition* **145**, 53–62 (2015).
82. Richter, A. & Courage, M. L. Comparing electronic and paper storybooks for preschoolers: attention, engagement, and recall. *J. Appl. Dev. Psychol.* **48**, 92–102 (2017).
83. Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* **44**, 978–990 (2012).
84. Goodman, N. D. & Stuhlmüller, A. *The Design and Implementation of Probabilistic Programming Languages* http://dippl.org/ (2014).

Acknowledgements
M.B. received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 749229. M.H.T. was funded by the National Science Foundation SBE Postdoctoral Research Fellowship grant no. 1911790. M.C.F. was supported by a Jacobs Foundation Advanced Research Fellowship and the Zhou Fund for Language and Cognition. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions
M.B., M.H.T. and M.C.F. conceptualized the study. M.M. collected the data. M.B. and M.H.T. implemented the models and analysed the data. M.B., M.H.T. and M.C.F. wrote the manuscript; all authors approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01145-1.

Correspondence and requests for materials should be addressed to M.B.

Peer review information *Nature Human Behaviour* thanks Lisa Pearl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.