Check for updates

RESEARCH ARTICLE SUMMARY

LANGUAGE LEARNING

The first 1000 days: Modeling language learning from individual children's everyday speech

Hadas Raviv*, Arkadii Tsyhanov, Kira Gousios, Aja Altenhof, Haocheng Wang, Berlin Chen, Ofri Raviv, Tal Rosenwein, Casey Lew-Williams, Liat Hasenfratz, Uri Hasson

Full article and list of author affiliations: https://doi.org/10.1126/science.aeh9311

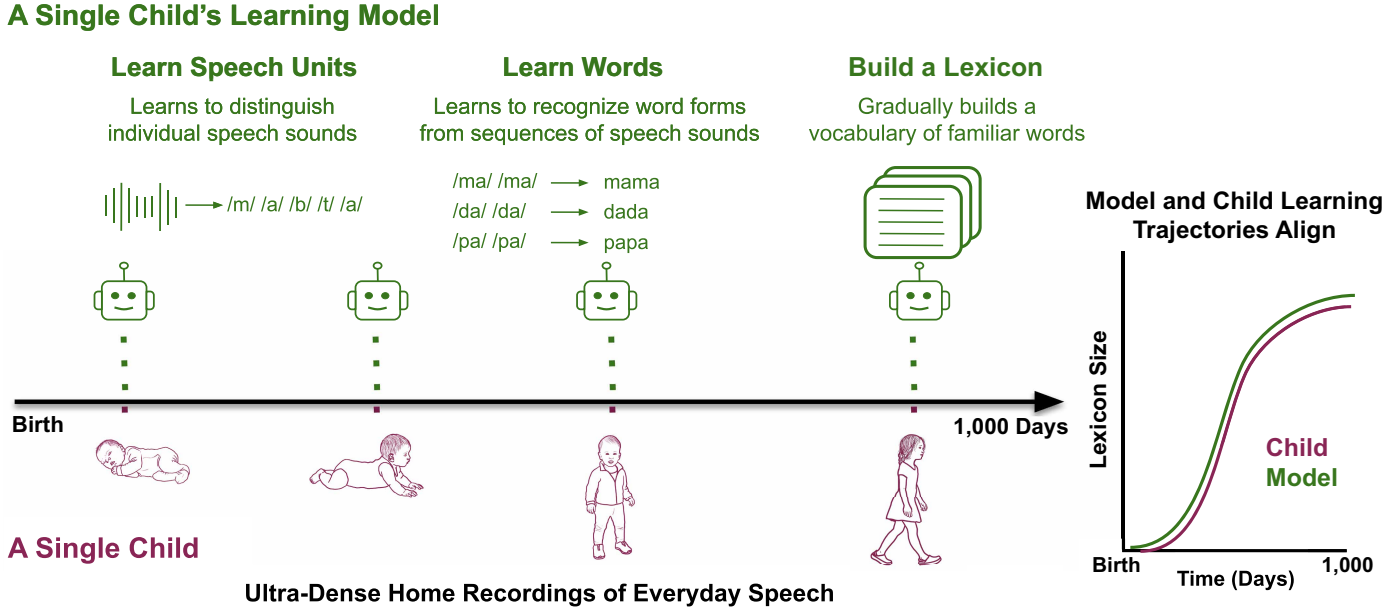
INTRODUCTION: Children learn to understand and produce speech from the language they hear every day, but how that everyday experience shapes learning remains an open question. Modern artificial intelligence (AI) language models show that language can be learned from data, yet they train on far more input than any child hears, using training protocols that are quite different from how children actually learn. Understanding how children learn language from everyday experience requires computational models built to learn the way children do, from the real-world input available to an individual child and on realistic developmental timescales.

RATIONALE: To test this idea, we used ultradense home recordings from the First 1000 Days project: thousands of hours of everyday speech collected for each of eight children across their first 1000 days of life. We built a separate learning model for each child, with no prior knowledge of speech or words. Each model learned only from the speech available to that child, processed day by day in the order it was originally experienced. Training followed a developmentally realistic schedule, with daily exposure gradually increasing as children spent more time awake and recent speech revisited at the end of each day to approximate the role that sleep may play in memory consolidation. We then tested whether the models learned distinct speech sounds and words and whether their learning pace tracked the developmental trajectory of the corresponding child.

RESULTS: Trained this way, the models learned to distinguish English speech sounds within roughly the first simulated year and went on to build vocabularies of thousands of words, with growth closely tracking each child's measured vocabulary. Two aspects of the experience were especially important. First, input density was critical: Models given only a fraction of a child's daily recordings learned far fewer words, likely because many everyday words are rare and require repeated exposure. Second, nightly replay was equally important: Models given little or no replay struggled to learn words, even when their daily input was rich. The same modeling framework worked across all eight children without modification. Within that framework, three internal model parameters governed the pace of learning, capturing individual differences and offering an interpretable account of how learning mechanisms interact with experience to support language acquisition.

CONCLUSION: Together, our results show that the everyday speech available to an individual child, combined with biologically inspired learning and memory processes, provides sufficient structure to support core aspects of language learning. More broadly, our approach offers a framework for translating theories of how children learn into computational models that can be tested directly against children's real-world development. □

*Corresponding author. Email: hr7485@princeton.edu Cite this article as H. Raviv *et al.*, *Science* **394**, eaeh9311 (2026). DOI: 10.1126/science.aeh9311



From a single child's everyday speech experience, a vocabulary emerged. Trained on individual children's day-by-day speech input across the first 1000 days, the model learned first to distinguish speech sounds and then words, gradually building a lexicon. From birth through toddlerhood, the model's vocabulary grew along a trajectory that closely matched each child's measured vocabulary growth. [Illustration credit: Z. Lavi-Hasson]

Downloaded from https://www.science.org at Princeton University on November 09, 2026

LANGUAGE LEARNING

The first 1000 days: Modeling language learning from individual children's everyday speech

Hadas Raviv[1]*, Arkadii Tsyhanov[2], Kira Gousios[1], Aja Altenhof[1], Haocheng Wang[1,3], Berlin Chen[1], Ofri Raviv[4], Tal Rosenwein[5], Casey Lew-Williams[3], Liat Hasenfratz[1], Uri Hasson[1,3]

Children learn to understand and speak from the speech around them, yet how that everyday input shapes learning remains unresolved. Large language models show that language is learnable from data, but their scale and training conditions differ sharply from a child's natural experience. We trained cognitively feasible, child-specific models on longitudinal home recordings from eight children across their first 1000 days of life. Each model learned incrementally from a single child's ordered speech input without prior linguistic knowledge, developing phoneme-like and word-form representations that tracked each child's learning pace. Learning was dependent on dense daily sampling and replay of recent input, indicating that both experience- and memory-related processes are necessary. These findings show that everyday experience contains sufficient structure for language learning and establish child-specific modeling for testing developmental theories.

Understanding how children acquire language from everyday experiences remains a long-standing interdisciplinary scientific challenge. Children learn from noisy, incomplete, and highly individualized input, yet they reliably converge on shared representations of speech and words without explicit instruction (*1*–*6*). A key unresolved question is whether naturalistic experience alone provides sufficient structure for early language learning (*4*–*7*) or whether successful acquisition depends on innate linguistic knowledge (*2*, *3*, *6*, *8*, *9*). This question arises at every level of language, from speech perception to word meanings to syntax, and extends beyond language to development more broadly (*9*). Here, we focused on one foundational component of language acquisition: whether the speech a child hears alone can support the emergence of language sounds and word-form representations on real developmental timescales (*10*).

Recent successes of large language models have shown that language can, in principle, be learned directly from input given sufficiently large training data (*11*–*13*). Yet the way these models learn differs radically from how children learn in at least three ways. First, they are trained on enormous datasets, often containing trillions of tokens (the word-like units that a model processes), far more than any child could encounter. Second, they rely on a predefined vocabulary of tens of thousands of such tokens, specified in advance rather than discovered from experience (*13*, *14*). Finally, their data-sampling and learning protocols bear little resemblance to the experience of any individual child (*15*–*17*). These differences raise a fundamental question: Can deep learning provide a viable computational framework for modeling human language acquisition? To address this question, we developed a unified framework that tested whether a child-specific model, trained on dense recordings from a single child's home over the first 1000 days of life, could support speech and word learning on developmentally relevant timescales.

Training such a model requires dense, continuous recordings of individual children. Existing ecologically valid datasets fall short, offering only sparse snapshots of experience (*18*–*23*). Capturing the full scope of an early linguistic environment is ethically and technically challenging. The First 1,000 Days (1kD) project (*17*) addressed this challenge, inspired by Deb Roy's pioneering Speechome project (*24*) while leveraging contemporary cloud infrastructure and automated machine-learning pipelines. It recorded 12 to 14 hours of naturalistic audiovisual data per day from 15 racially and economically diverse US families across each child's first 1000 days of life. In the present study, we focused on isolated speech derived from the recordings of eight children within this corpus.

We trained a child-specific, cognitively feasible learning model on the daily speech input of one child at a time and replicated this process across eight children, each from a different home environment. The model's weights were randomly initialized with no built-in linguistic knowledge. We simulated learning on developmentally realistic timescales and evaluated each model by comparing its learning dynamics with the developmental trajectory of the corresponding child. The training protocol was designed to model development under ecologically feasible constraints: Learning proceeded continuously day by day, with replay-like reprocessing of prior input inspired by memory consolidation (*25*, *26*), and followed a curriculum aligned with the stages of early language development (*27*). The model was trained on isolated, resynthesized speech rather than raw sensory input, leaving open the question of how learning proceeds from noisier real-world input such as background noise, overlapping conversations, and multiple speakers.

We found that each model acquired speech-unit representations and word-level abilities that children typically develop within the first 2 years of life (*27*–*34*). Speech-unit representations emerged by learning from continuous, household-specific speech input and became increasingly aligned with phonetic structure over development. These representations supported the emergence of thousands of word forms, with lexical growth tracking each child's developmental trajectory as measured by standardized behavioral assessments [MacArthur-Bates Communicative Development Inventories (CDIs) (*35*)]. Critically, the same architecture generalized across all eight children. Differences in learning rate were captured by variation in three internal model parameters, providing an interpretable account of how learning mechanisms interact with experience to support language acquisition. Reproducing individual children's trajectories depended on the combination of rich daytime input and nighttime replay-like reprocessing, indicating that the structured input of a single child's life can support early language learning when paired with consolidation-like processes.

More broadly, this approach grounds an artificial learner in both the content and temporal structure of a child's natural experience. Together, the 1kD recordings and learning model provide an extensible computational test bed for identifying which additional mechanisms and modalities are needed to better account for human language learning under real-world developmental constraints.

Results

Ultradense 1000-day recording of language development at home

Data collection and processing: With human-subject approval (Princeton University institutional review board; protocol no. 13884), a data-safety monitoring board, and parental consent, we installed audiovisual recording systems in 15 racially and economically diverse US homes to capture children's daily experience across the first 1000 days of life (Fig. 1A). Recordings were conducted for 12 to 14 hours per day in key household spaces using one or two cameras with built-in microphones and one high-fidelity external microphone per room. We processed the resulting ~1.2 million hours of data using a scalable multimodal

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. [2]Sigma Software, New York, NY, USA. [3]Department of Psychology, Princeton University, Princeton, NJ, USA. [4]Independent Consultant, Princeton, NJ, USA. [5]Independent Consultant, Tel-Aviv, Israel.
*Corresponding author. Email: hr7485@princeton.edu

Downloaded from https://www.science.org at Princeton University on November 09, 2026

analysis pipeline designed to detect speech near each child (*17*). Here, we focused on the eight homes for which recording, preprocessing, and modeling were completed. This subset preserved the racial and socioeconomic diversity of the broader cohort and included four female participants. These families were recorded for a median of 960 days (range: 495 to 990 days), yielding ~1900 to 6600 hours of naturalistic audio per child and corpora of ~6 million to 27 million automatically transcribed words (Fig. 1B; all children's names are pseudonyms). Segmentation by awake-time schedule (see materials and methods) produced developmental timelines spanning 14 to 34 months across households.

Because the naturalistic far-field recordings included background noise and reverberation, we simplified the input by resynthesizing clean speech-only signals from the transcripts using a text-to-speech model (*36*). This enabled us to test how children learn from the speech present in their environment while postponing the challenge of modeling full acoustic complexity for future work.

Quantifying children's word acquisition: In addition to the continuous audiovisual recordings, we collected monthly caregiver-reported measures of children's language development using the MacArthur-Bates CDIs (*35*). From 8 to 16 months, we administered the Words and Gestures form to assess comprehension and production. At 16 months, in accordance with standard practice, we transitioned to the Words and Sentences (Toddler) form to track speech production through the end of the study.

Word-learning trajectories varied noticeably among children, with the age at which half of the CDI words were acquired ranging from 12 to 22 months (Fig. 2). To derive these trajectories, we defined the age of acquisition for each lexical item as the first month in which the caregiver reported that the child knew the word. The cumulative number of learned words as a function of age is shown as purple circles (comprehension) and orange squares (production) in Fig. 2A. We fit a sigmoid function to each trajectory (see materials and methods). For each fit, we estimated two parameters: time to learn, the age at which 50% of the words had been acquired, and steepness, which reflects the rate of learning. The fitted parameters for comprehension trajectories across the eight children are shown in Fig. 2, B and C. These analyses reveal interindividual differences in word-learning trajectories that are visible only at the level of densely sampled individual children. Notably, variability in time

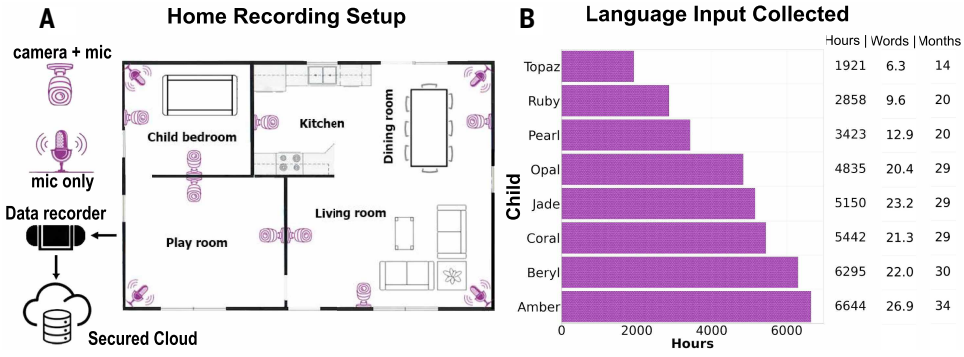


Fig. 1. Overview of in-home recording setup and data collection. (**A**) Illustrative layout of recording equipment across key rooms in a participating home (not an actual floor plan). Audio and video streams were routed to a local data recorder and uploaded to a secured cloud server. (**B**) Linguistic input collected for each of the eight children, identified by pseudonyms. Bars show the total number of speech recording hours per child. The adjacent columns report these hours, the total number of words (in millions), and the effective number of recording months.

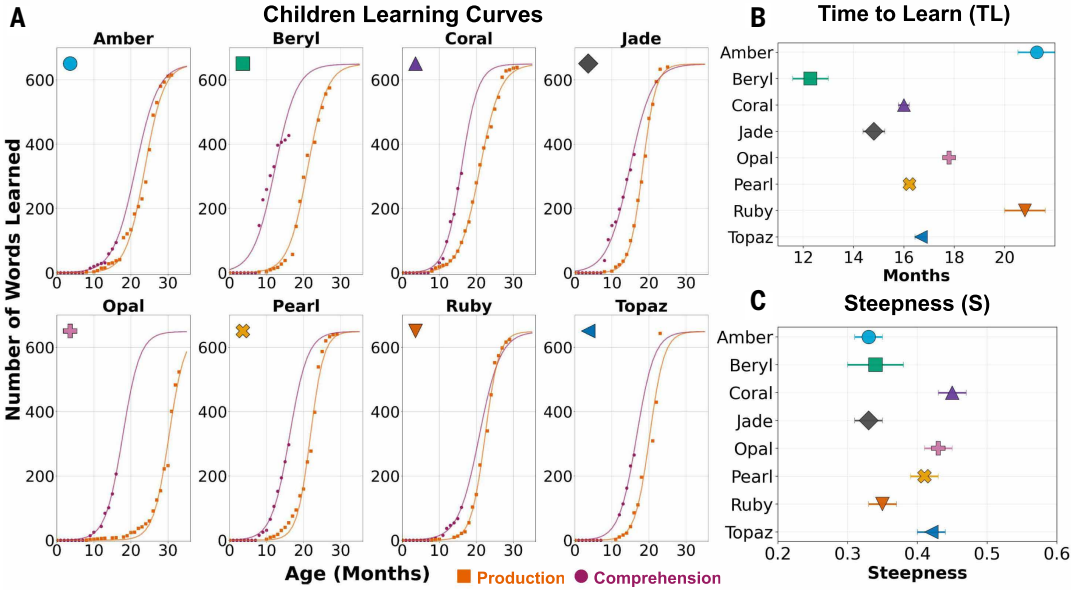


Fig. 2. Word-learning trajectories derived from parental reports. (**A**) Cumulative number of words learned as a function of age for comprehension (purple circles) and production (orange squares), based on parental responses to the CDIs. Solid lines show fitted sigmoid curves for each of the eight children. Comprehension was measured with the CDI Words and Gestures form from 8 to 16 months, and comprehension trajectories beyond 16 months reflect extrapolations from the fitted curves. (**B**) Fitted values of time to learn—the age at which 50% of the words were acquired—for the comprehension curves in (A). (**C**) Fitted values of steepness—the steepness of the curve, which reflects learning rate—for the same curves. Error bars indicate approximate 95% confidence intervals for the fitted sigmoid parameters. Marker shape and color identify individual children and are used consistently across the figures.

Downloaded from https://www.science.org at Princeton University on November 09, 2026

to learn (Fig. 2B) exceeded variability in steepness (Fig. 2C), suggesting larger differences in when children reached mid-vocabulary than in how quickly vocabulary grew once acquisition began.

Designing a cognitively feasible child-specific learning model

The ultradense recordings enabled us to train a self-supervised learning model that simulates how everyday experience interacts with internal learning processes over time, supporting the emergence of phoneme and word-form representations. Starting from random weights, the model learned solely from each child's speech input, without prior linguistic knowledge, meaning that any phonetic or lexical structure it acquired originated in that child's experience. Learning followed a cognitively motivated training protocol that incorporated day-by-day learning, a developmental curriculum, and replay-like memory consolidation.

Learning model architecture: Our cognitively motivated learning model acquired language directly from continuous speech in a fully self-supervised manner. It used an encoder-decoder architecture that separates learning into two stages. First, the encoder discretized the continuous speech stream into 20-ms units, analogous to phonemic or subphonemic elements (*30*, *37*–*40*). Second, once the encoder stabilized, the decoder learned sequential regularities by predicting the next units based on prior context. As prediction improved, the system treated reliably predicted word forms as learned tokens and stored them in an incrementally updated dictionary (Fig. 3A). The encoder is based on DINO-SR (*41*), which learns cluster-centered representations of auditory space via a data-driven, unsupervised objective. The decoder is a speech-tailored transformer adapted from Kharitonov *et al.* (*42*). Full architectural details are provided in the materials and methods.

Training protocol: The cognitively grounded training protocol had four key features:

1) The model's learning trajectory was explicitly mapped onto a real developmental trajectory (days, months, and years). The model consumed audio in the chronological order in which it was recorded, thereby preserving the temporal structure of the input, unlike conventional training regimes that often rely on shuffled data. The basic temporal unit of training was one developmental day (Fig. 3B). On each training day, the model received the number of words the corresponding child heard during a typical day at that age. Because children remain awake for longer periods as they grow, this daily amount increased over development. This was computed by multiplying the words-per-hour rate from that child's household recordings by the number of waking hours estimated for that age (see Training protocol, point 4).

2) An important internal model parameter was the error-accumulation window: the duration in minutes over which prediction errors accumulated before a single weight update. Because speech exposure unfolds over measurable periods of time, this window determined how frequently the model updated after each fixed duration of exposure. We fitted the error-accumulation window separately for each child and applied it during both daytime learning and nightly replay. This mechanism may have a parallel in the brain, where higher-order regions accumulate information over minutes of speech rather than moment to moment (*43*).

3) We incorporated a biologically inspired replay mechanism (Fig. 3B). Standard deep learning organizes training into epochs, each defined as a full pass over the dataset, with examples shuffled across iterations. In our setting, this would amount to a child repeatedly reexperiencing their first 1000 days of input in random order, which is developmentally implausible. Instead, at the end of each training day, we augmented that day's training input with speech segments replayed from earlier days, sampled with a probability that decayed with temporal distance from the current day (see materials and methods). Recent input was therefore replayed most often, with older input replayed progressively less often, analogous to the sleep-dependent replay implicated in memory consolidation (*25*, *26*).

4) Learning followed a developmental curriculum that approximated changes in children's real-world experiences, implemented in two complementary ways. First, we progressively increased the amount of daily speech input available to the model to reflect developmental changes in children's waking time (Fig. 3C). This waking schedule was held constant across children. Second, we staged learning across the encoder and decoder (Fig. 3D). Because stable speech-unit representations corresponding to phonemic or subphonemic elements are plausibly a prerequisite for learning longer-range regularities such as word forms, we modulated the decoder's effective learning rate using a sigmoid schedule. This attenuated decoder updates early in development while preserving continuous training, so that sequential patterns were learned effectively only after the encoder had stabilized its speech-unit representations. This schedule was preset rather than learned; future work could replace it with an internally derived gating signal.

The first feature was determined by each child's measured speech input, whereas the other three were model-internal parameters fitted separately for each child:

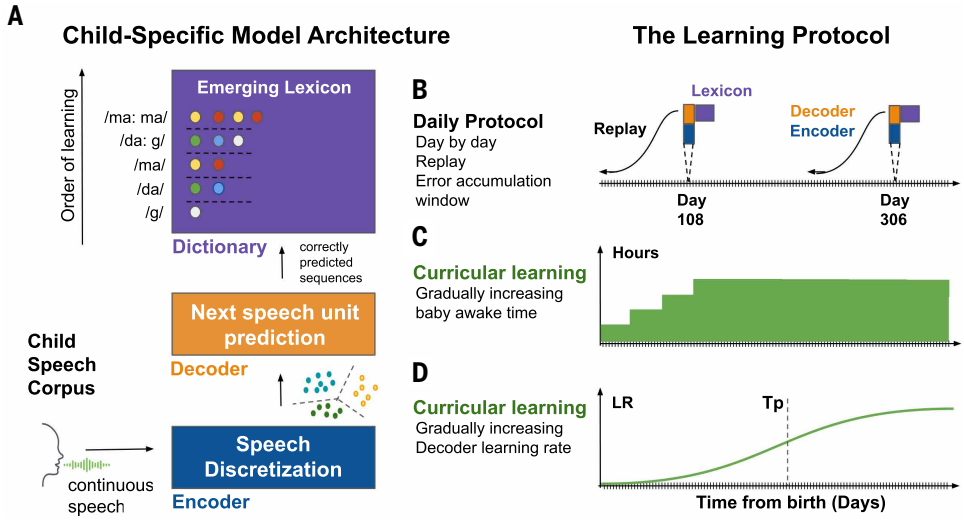


Fig. 3. Overview of the model architecture and training protocol. (**A**) The model comprised three modules. The encoder was trained on continuous speech from the child speech corpus and learned how to map 20-ms windows of continuous speech to discrete speech units. The decoder received discrete units and their durations and learned to predict the next unit. The lexicon aggregated reliably predicted sequences over time. For example, the sequences shown as "/ma: ma/" and "/da: g/" correspond to the words "mama" and "dog," respectively. (**B** to **D**) Cognitively motivated training protocol. (B) shows daily learning from the amount of speech estimated directly from the recorded home data, combined with replay, implemented as daily resampling of past experience with a probability that decayed with temporal distance. The error-accumulation window is the duration in minutes over which prediction errors accumulated before a single weight update. (C) presents curriculum learning over input quantity, introduced by gradually increasing the children's waking hours. (D) shows curriculum learning over decoder learning rate (LR), implemented as a sigmoidal schedule, with Tp marking the midpoint of the transition.

Downloaded from https://www.science.org at Princeton University on November 09, 2026

error-accumulation window, replay, and the decoder's learning-rate schedule (Fig. 3, B to D; see materials and methods). Together, the measured input and the fitted internal parameters accounted for individual differences in learning rates across children, allowing the model to simulate how experience and internal learning mechanisms interact over development on realistic timescales.

Evaluating the model's learning trajectories: A central challenge in this work was determining which phonemes and word forms the learning model acquired at any given point in development. Given the time-dependent nature of child learning, in which linguistic elements are encountered intermittently and at varying frequencies, we used a weekly evaluation framework to assess the model's development. Specifically, we evaluated the model's learning at the end of each week using data from the following week that had not been used for training. This testing protocol mirrors real-world development: Both the model and the child learn through the ordered sequence of experiences accumulated up to that point, and their knowledge state is assessed on subsequent, previously unseen input.

Training a learning model based on the linguistic input of a single child

We first present the full training, evaluation, and analysis workflow for a single household, focusing on one child (Coral, a pseudonym) as an illustrative case study. The learning model was trained on all child-centered speech from Coral's home, following the protocol described above, with random initialization and no prior linguistic knowledge. Because the current experiments use clean, resynthesized speech, the first 2 months of developmental time (days 0 to 60) were treated as a placeholder for future work on separating speech from background sounds in naturalistic audio. We first characterize the development of Coral's model speech-unit representations and then analyze the emergence of word-form learning over time. We then extend the same analyses across all eight models to evaluate robustness and generalization across households.

Quantifying phoneme acquisition in learned speech units: After 24 months of training, Coral's model mapped learned speech units onto 38 of the 39 English phonemes (Fig. 4A). The encoder outputs 512 discrete speech-unit labels at 20-ms resolution, which allowed us to quantify unit-phoneme associations by computing, for each unit and each phoneme, the probability that the unit was assigned to that phoneme during evaluation (see materials and methods). Across 38 phonemes, each was supported by at least one highly selective unit (association of >0.5), whereas most units showed weak associations with other phonemes. For each phoneme, the most selective unit had a mean association of 0.91 (range: 0.53 to 1.00); the only substantially weaker case was /ZH/, which was rare in the data. These results indicate that the learned units captured fine-grained phonemic structure, with multiple units representing individual phonemes.

Most English phonemes became reliably discriminable by 10 to 12 months of training (Fig. 4B). To track learning, we computed the unit-phoneme association matrix at the end of each training week and defined a phoneme as learned when at least 30% of its instances were associated with selective units (association of >0.5; see materials and methods). Performance rose rapidly between months 5 and 9, reached a high level by month 9, and largely stabilized by month 16. By ~16 months, 38 of 39 phonemes had been learned. Supplementary analyses further showed that these mappings retained a many-to-many structure, preserving fine-grained detail while increasingly consolidating around dominant units for each phoneme over development (fig. S1, A to D).

Quantifying word-form learning: The decoder learned to predict each upcoming speech unit based on prior context. We counted a word as learned once the model reliably predicted the sequence of speech units within that word's boundaries (*44*, *45*). Specifically, we used a transcript-aligned window spanning at least 90% of the word duration and scored a word occurrence as correctly predicted when the model's confidence across the speech units was significantly higher than a random baseline and remained stable or improved across the sequence (see materials and methods). We defined a word type as learned at time t once it had been correctly predicted in at least $N = 15$ occurrences across different contexts.

Coral's model acquired word forms at a pace that closely matched their observed development, as measured by monthly CDI assessments (Fig. 5, A to C). Comprehension was assessed in two phases. Between 8 and 16 months, caregivers completed the CDI comprehension form, which was used directly to estimate Coral's knowledge of 649 curated CDI words (solid purple bars in Fig. 5A; excluding phrases, see materials and methods). Because comprehension data collection ended after 16 months, we estimated subsequent growth for the same set of words through 24 months using a sigmoid fit (striped purple bars in Fig. 5A). At 12 months, Coral was reported to understand 97 words (model fit: 125); by 24 months, predicted comprehension reached 632 (Fig. 5A). Like Coral, the model learned nearly all CDI words along a comparable trajectory, with similar growth rates and learning dynamics (Fig. 5B). An analysis of model confidence revealed two phases of word learning (fig. S2): Prediction confidence increased from ~9 months and stabilized at high levels by ~17 months. Sensitivity analyses across the parameters of the word-form learning rule indicated that the captured developmental dynamics were robust across various conditions (fig. S3).

This alignment depended on the ultra-dense 1kD recordings (Fig. 5C). Training on different fractions of the daily data revealed a nonlinear relationship between recording density and the number of words learned. Using only 10% of daily data (about 1 hour of daily recordings), the model learned no CDI words by 24 months. At 30% (around 2.6 hours daily), it learned only 58 of 649 words, roughly 9%. With 70% of the daily recordings, the model learned roughly 90% of the CDI words

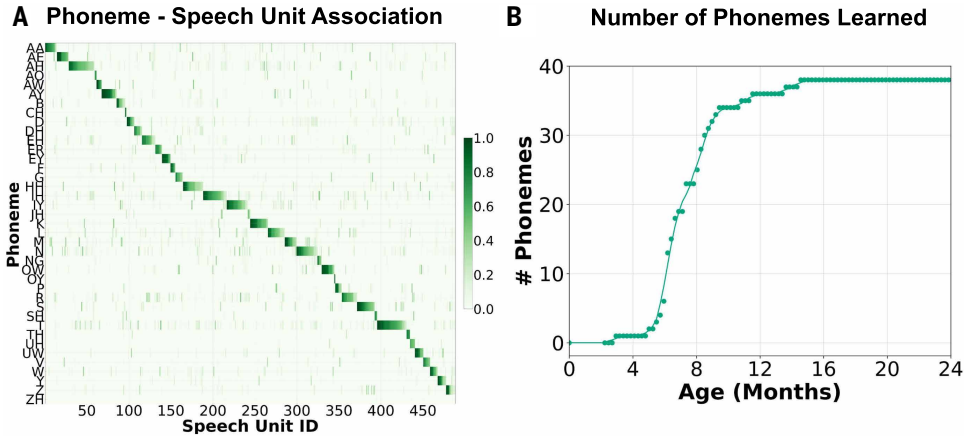


Fig. 4. Phoneme-unit structure and phoneme acquisition over time. (**A**) Phoneme-speech unit association matrix at 24 months for Coral's model. Color indicates the proportion of each speech unit's tokens that are aligned to a given phoneme, with units ordered by their dominant phoneme. (**B**) Cumulative number of learned phonemes over development, using the phoneme-coverage criterion.

Downloaded from https://www.science.org at Princeton University on November 09, 2026

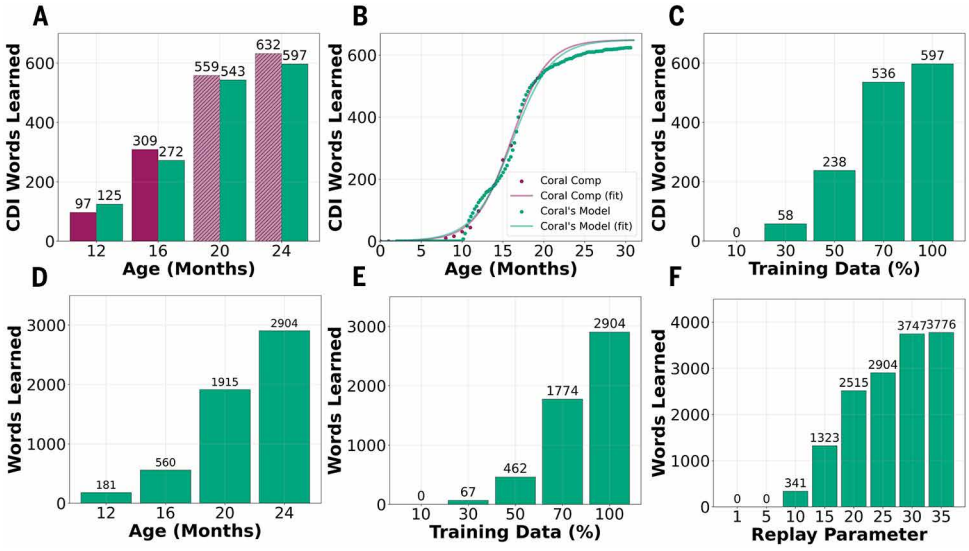


Fig. 5. Word-learning trajectories from a single-child home environment (Coral). (**A** to **C**) CDI analysis. (A) shows the number of CDI words known by Coral from monthly CDI comprehension reports (solid purple bars) and extrapolated estimates (hatched purple bars), alongside the number of CDI words learned by the model (green) at 12, 16, 20, and 24 months. (B) presents the full CDI word-learning trajectories for Coral (purple) and the model (green), showing observed values and sigmoid fits. (C) shows the effect of daily input quantity: CDI words learned at 24 months as a function of the percentage of available daily data used for training. (**D** to **F**) Open-vocabulary analysis. (D) shows cumulative learned lemmas over the development period. (E) presents the effect of daily input quantity: cumulative learned lemmas at 24 months as a function of the percentage of available daily data used for training. (F) shows the effect of replay on word learning: learned lemmas at 24 months as a function of the replay parameter.

learned using the full dataset. Full alignment between the trajectories of Coral and the model required all recordings, likely reflecting the heavy-tailed distribution of language input. Ultradense recordings were therefore essential for capturing the statistics of children's real-world language experiences.

The CDI list captures only part of the words Coral heard, so we estimated the model's capacity to learn other word forms in the home environment. We counted word types at the lemma level, collapsing variants (such as eat, eating, ate) into a single entry; stratifying by part of speech yields the same qualitative pattern. Figure 5D shows the cumulative number of learned lemmas at 24 months. Because open-vocabulary size is not fixed, we report raw cumulative counts rather than fitting a parametric growth curve. As in the CDI analysis, the number of learned word forms increased steadily with age, reaching thousands of lemmas by 24 months. Vocabulary growth also depended strongly on recording density (Fig. 5E): Training on 50% of the daily data yielded only about 16% of the lemma vocabulary achieved with the full 8 hours per day by 24 months. This pattern suggests that denser sampling enables the model to acquire the long tail of low-frequency words not captured in the CDI list.

Replay strongly shaped learning: Although Coral's language environment was rich, with ultradense daylong recordings totaling roughly 21 million words, a single pass over this input was insufficient for learning (Fig. 5F, one exposure). At the end of each training day, we replayed recent input by resampling prior events in amounts matched to that day's awake hours, using an inverse-weighting function that favored more recent events (see materials and methods). This replay was critical for learning (Fig. 5F). With little or no replay (one to five exposures), the model failed to learn words by 24 months, consistent with a role for replay-like mechanisms in early learning (*46*). Replaying each daily event roughly 25 to 30 times was sufficient, with additional exposures providing little benefit. Replay did not substantially alter the lexical distribution of the training input (fig. S4).

We also examined the error-accumulation window size (see materials and methods). Word learning at 24 months increased as the size of the error-accumulation window was shortened from 1 hour to 10 min, declined slightly at 8 min, and failed by 4 min (fig. S5). Language acquisition, therefore, depended on both the amount of experience and the frequency with which that experience was processed. Across the error-accumulation window, replay, and curriculum parameters, the full learning trajectories (fig. S6) showed that internal model parameters substantially shaped each child's developmental learning trajectory, whereas the learning rule itself remained fixed.

To qualitatively examine the emerging word-form representations, we analyzed word embeddings learned at 24 months. For each CDI category, we projected these embeddings and their category centroids into two dimensions (see materials and methods). The resulting spaces were highly structured (fig. S7): Words formed local clusters that largely reflected acoustic similarity among word forms (for example, star-stone). Acoustic similarity did not fully account for this organization, however. Some categories showed hints of category-level structure (such as sun-rain-wind), suggesting that early word-form representations may begin to reflect aspects of word meaning.

Generalization across children: A central goal of developmental psychology is to explain both shared developmental mechanisms and meaningful individual differences. So far, we have shown that the learning model can simulate a single child's developmental trajectory in a real-life setting using speech derived from that child's everyday input. We next tested whether the framework generalized across children growing up in distinct home environments. Specifically, we asked whether (i) the same learning architecture could support learning from the distinct linguistic input available to seven additional children and (ii) the framework could account for meaningful individual differences in vocabulary acquisition and learning rates across all eight children.

To investigate these questions, we trained two models for each of the seven additional children, whose everyday experiences were recorded and processed using the same protocol. The first model used Coral's internal learning parameters to test whether a common learning mechanism could produce similar developmental patterns across different home environments. The second model used three internal learning parameters fitted separately for each child, testing whether variation in the model's internal learning dynamics could account for child-specific developmental trajectories.

As with Coral's model, we quantified phoneme and word-form learning over time and extracted summary parameters by fitting learning curves to each child and to the corresponding model. For five out of the eight children, the ultradense recordings provided sufficient speech input to support training for ~30 months (Fig. 1B). For the remaining three children, the available isolated-speech input was more limited, totaling 1900 to 3500 hours, corresponding to ~14 to 20 months of training data. For these children, we extrapolated the model's

Downloaded from https://www.science.org at Princeton University on November 09, 2026

word-learning trajectory over the remaining months using the same sigmoid-based procedure used to estimate children's comprehension beyond 16 months of age (Fig. 2A and materials and methods).

Phoneme learning was robust and generalized across all eight children (fig. S8). Each model was trained on speech derived from a single child's household and acquired an average of 37.5 phonemes (range: 36 to 38) within a similar developmental window of ~12 to 15 months. Topaz's model acquired 36 phonemes, likely reflecting the smaller amount of training data available for this child.

Word learning was robust to input variation across households (Fig. 6A). We instantiated the same model architecture with Coral's internal learning parameters and trained a separate model from scratch on speech from each of the eight households. Every model acquired word forms from its corresponding household input, and differences in input produced only modest shifts in learning timing (Fig. 6A). The estimated age at which the models acquired 50% of the CDI vocabulary ranged from 16 to 18 months (Fig. 6B). The learning framework was therefore robust and reliably supported word-form acquisition across distinct home environments.

Individual differences in the pace of word learning were captured by varying three internal learning parameters across children (Fig. 6, C to E). Children acquired half of the CDI vocabulary between ~12 and 22 months (x axis of Fig. 6B), a substantially wider range than the 16 to 18 months produced by differences in household input when model parameters were held fixed (y axis in Fig. 6B). Variation across households was therefore insufficient to explain the observed differences in developmental timing. We fitted three parameters separately for each child: the learning-rate schedule, the error-accumulation window, and the amount of replay. Each had a distinct cognitive interpretation and was varied within a constrained range, effectively pairing each home environment with a learner characterized by distinct internal learning dynamics. The individually fitted models closely tracked the corresponding children's developmental trajectories and captured variation in the timing of word learning (Fig. 6C and fig. S9). In particular, they reproduced the ~12- to 22-month range in the age at which children acquired 50% of the CDI vocabulary (Figs. 2B and 6D). Variation in learning-curve steepness was captured only partially (Fig. 6E): The models generally began acquiring word forms later than the children and subsequently progressed more rapidly. Sensitivity analyses of the word-form learning criterion showed robust acquisition across speech-unit prediction-confidence thresholds ranging from 0.05 to 0.5 for all child-specific models (fig. S10). Even at a stringent threshold of 0.4,

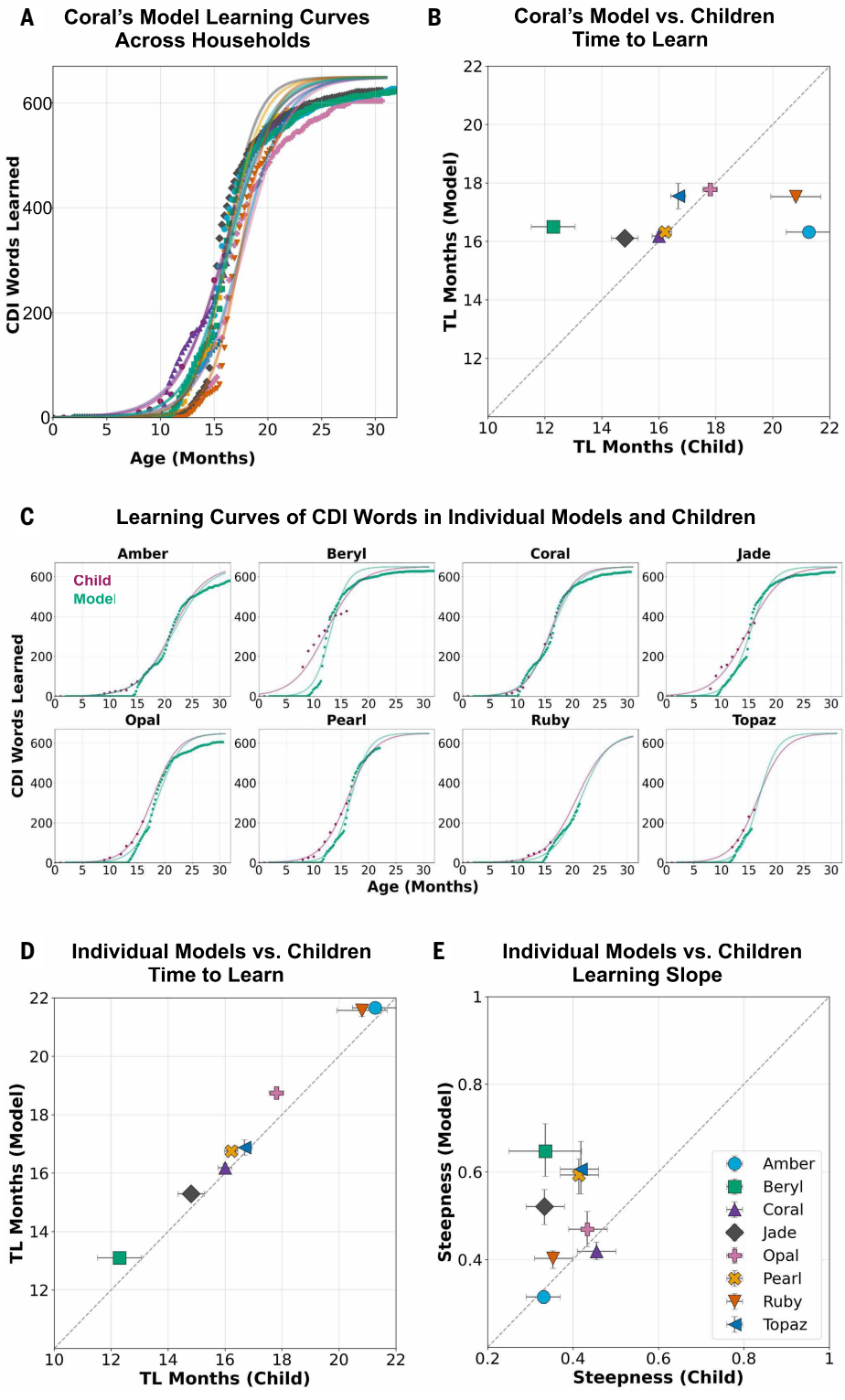


Fig. 6. Learning trajectories across different home environments. (**A**) Word-learning trajectories for models trained on each child's input using the internal learning parameters of Coral's model. Points show the cumulative number of CDI words learned over development, and lines show sigmoid fits to the full trajectories. (**B**) Time to learn 50% of the CDI vocabulary (TL) for these fixed-parameter models plotted against the corresponding child's fitted TL. The dashed line indicates identity. (**C**) Word-learning trajectories for models whose internal learning parameters were fitted separately for each child. Children's trajectories are shown in purple and model trajectories in green. (**D**) Time to learn 50% of the CDI words and (**E**) learning-curve steepness. Each point represents one child-model pair, error bars indicate approximate 95% confidence intervals for the fitted sigmoid parameters, and dashed lines indicate identity. Marker shape and color follow those used in Fig. 2.

Downloaded from https://www.science.org at Princeton University on November 09, 2026

every model with sufficient training time acquired the full CDI vocabulary by ~20 to 30 months, indicating that the models ultimately developed high-confidence predictions for all words in the independent evaluation sets. Finally, as observed for Coral's model (Fig. 5D), the CDI vocabulary represented only a subset of the rich lexicon acquired by each individual model (fig. S11).

Discussion

Children grow up in complex, dynamic, and highly individualized environments, yet converge on broadly shared linguistic systems. To examine how this convergence can arise from real-world experience, we paired ultradense longitudinal recordings of children's everyday speech environments with child-specific models trained on each child's unique speech input. The models were initialized with random weights without built-in knowledge of speech or language structure and were trained using a cognitively feasible learning protocol. When the same model architecture and internal parameters were applied separately to recordings from each of the eight households, every model acquired speech-unit representations that were increasingly aligned with English phonemes and developed a large lexicon of word forms, despite substantial variation in input across homes. Allowing three internal parameters (learning-rate schedule, error-accumulation window, and amount of replay) to vary across children further enabled the models to capture individual differences in learning pace. Together, these findings show that everyday speech contains sufficient structure to support early speech and word-form learning when coupled with effective learning mechanisms and offer a framework for explaining both shared developmental patterns and meaningful individual differences.

The model's learning architecture and training protocol provide a mechanistic bridge between early speech processing and lexicon formation. Phoneme-like representations emerged as regularities across fine-grained 20-ms discrete speech units. Critically, these units are themselves learned categories of acoustically similar segments, discovered directly from environmental speech by an encoder designed to identify reliable structure in the continuous acoustic signal. Word-form knowledge, in turn, emerged in the subsequent prediction module (decoder) as increasingly predictable sequences of these discrete units within word boundaries. Both phoneme-like and word-form representations were acquired without explicit language priors or external supervision. This account aligns with mechanistic implementations of self-supervised speech learning (*37*, *38*, *40*) and with behavioral evidence of gradual tuning to native-language speech units in the first year of life (*27*, *30*, *33*). The timescales over which phoneme- and word-form representations developed closely matched prior reports (*30*) and those observed in our focal children. Importantly, the proposed staged architecture is one possible mechanistic account of learning at both the speech and lexicon levels and should not be read as claiming that phoneme and word learning are strictly sequential in children (*47*–*49*). The same framework could test alternative interactive accounts, for example, by allowing emerging lexical representations to feed back onto speech-unit representations.

Development unfolds in rich, multidimensional environments that sparse measurements do not capture. The 1kD recordings provide a dense view of children's everyday experiences, with a temporal resolution that extends well beyond traditional laboratory sampling or brief naturalistic observations. These ultradense recordings were essential for training the model: Learning performance degraded substantially when daily speech input fell below 8 hours. Our findings suggest that early language acquisition depends on two related factors: a rich linguistic environment that supports early speech learning and boosts vocabulary growth and sufficient long-tail statistics that enable acquisition of infrequently used words, which are typically learned slowly (*50*). More broadly, modeling naturalistic development requires dense, continuous sampling of language input as it unfolds over minutes and hours and accumulates across weeks, months, and years.

The speech environment was rich, but richness alone was not sufficient. Without daily replay of recent input, the model failed to form stable speech units or accumulate lexical structure over developmental timescales. Thus, experience alone was insufficient to support learning in our model; it had to be augmented by memory of past input. Revisiting past data is common in deep learning, but standard training does so in a way not cognitively feasible for a child: Models repeatedly cycle through the entire dataset in reshuffled order, an experience no child encounters. Our replay instead resampled recent input at the end of each day, weighted toward the most recent experiences and loosely modeled after sleep-dependent memory consolidation. We view this as a hypothesis rather than a finding: Children, like the model, may rely on a replay-like process that reuses recent linguistic experience without full reexposure. Two lines of evidence support this hypothesis. First, infants sleep extensively during the period studied here (*51*), and infant sleep has been linked to memory consolidation and word learning (*46*, *52*–*54*). Second, replay of recent experience was demonstrated in rodents and humans, with hippocampal sequences replaying in temporal order during sleep in rodents (*55*) and sequential replay observed during rest in humans (*56*). Neither line of evidence specifically establishes the replay of speech sequences. However, within the framework presented here, replay offers an open and testable candidate mechanism linking perception, memory, and learning during early cognitive development (*57*).

Our implementation of replay makes several simplifying assumptions that future work should examine directly. The recency-decayed resampling function used here was chosen as a reasonable first approximation, rather than one independently established by prior data. Indeed, the broader literature on replay prioritization remains unsettled, with some findings linking replay to recency and others to reward, surprise, or infrequency of prior exposure (*25*, *26*). Our implementation also assumes that replay is content neutral, with no role for emotional salience or surprise, as has been shown to bias replay in other memory domains (*58*, *59*). Finally, the model replays retrieved samples without error or degradation, an assumption that is unlikely to hold for biological memory retrieval, which is reconstructive and imperfect (*43*, *60*). Testing alternative decay functions, content-based replay biases, and noisy or error-prone retrieval against the suggested baseline is an important direction for future work.

Our mechanistic framework provides a data-driven benchmark for developing and evaluating theories of cognitive development. By integrating structured learning architectures with rich, ecologically valid input, researchers can model how learning unfolds over individual developmental timelines in real-world contexts. Constructs typically studied in brief, controlled laboratory tasks can thus be reformulated as explicit computational hypotheses about how naturalistic input shapes developmental outcomes. The framework's emphasis on mechanistic specificity is particularly important: It requires theories to be instantiated as algorithms that learn under the temporal and statistical constraints that children face in everyday life while enabling direct comparisons of their explanatory value against alternative accounts (*61*, *62*). More broadly, the framework does not commit to a single mechanism of language learning; it provides a shared test bed for comparing alternative computational hypotheses under the same child-specific input, temporal structure, and observed developmental outcomes. This includes testing whether individual differences in learning trajectories, such as onset timing, steepness, and word learning order, are better explained by differences in input, internal learning parameters, or omitted modalities (for example, vision). In this sense, mismatches between model and child trajectories are not merely failures of fit but diagnostic

Downloaded from https://www.science.org at Princeton University on November 09, 2026

constraints that identify which mechanisms future models must refine or add. Finally, investigating how the model's internal predictive capacity evolves over time raises questions about how we define and measure learning and suggests that subtler behavioral tests of children's vocabularies could more effectively capture the nuanced changes in their developing knowledge.

Several limitations delineate important directions for future work. First, the current model was trained on clean, resynthesized speech, which necessarily simplifies the acoustic complexity of everyday input. Resynthesis preserves lexical content but removes the original prosody, speaker identity, and register-specific features of natural speech. Furthermore, the acoustic and social properties of child-directed speech input have been implicated in infant speech processing and language learning (*63*–*65*). Modeling their contribution will require tools that can separate the original speech signal from background noise and overlapping talkers. Second, the architecture was designed to be cognitively, rather than neurally, plausible. Transformers process context in parallel rather than through the recurrent, serial dynamics characteristic of cortical circuits, and they are trained with backpropagation, a learning rule whose exact biological implementation remains debated (*66*). Even so, transformer-based representations have been found to align with human neural activity during language processing (*67*), suggesting that some of the structure that they capture may be representationally relevant despite this architectural mismatch. Future work could build on this framework by incorporating more biologically plausible learning mechanisms. Third, the present framework omits multimodal information and social context, both of which are central to children's everyday learning (*68*). The 1kD dataset includes synchronized video alongside audio, making a multimodal extension of the present model a natural next step (*17*). Future embodied and interactive extensions could further incorporate active exploration and action selection, linking perception, action, and language learning more closely to children's real-world experiences. Finally, extending the framework beyond early sound and word-form learning to syntactic and pragmatic learning will test whether similar computational principles scale to higher levels of linguistic structure (*69*, *70*). The current architecture and the 1kD dataset were designed with these extensions in mind, and the present model is best understood as a first concrete step in this longer research program.

This research program also raises a broader question for the field: whether developmental science should aim to build increasingly accurate models of how individual children learn. We believe that the scientific and clinical value of this approach, including the possibility of identifying atypical learning trajectories earlier than population-level screening allows, justifies pursuing it, provided it remains responsibly governed. Accordingly, access to the underlying data and models should remain restricted for approved research use and subject to controlled access procedures and privacy protections. This work explicitly aims to test developmental hypotheses, not to construct general-purpose, humanlike systems. Maintaining such boundaries is an ongoing responsibility as models become more capable.

The 1kD dataset together with the child-specific learning model provides a mechanistic account of how rich environmental input interacts with internal learning mechanisms to support early language development in individual children. Furthermore, it demonstrates that foundational aspects of language—phoneme-like and word-form–like representations—can emerge through adaptive direct fit to the rich structure of young children's natural environments (*71*). More broadly, this modeling approach aligns with views of development as an emergent product of multiscale interactions unfolding over time, rather than the result of fixed, prespecified representations (*72*). Together, this framework offers a concrete foundation for the next generation of empirically grounded, ecologically robust theories of language acquisition.

Materials and methods are available in the supplementary materials.

REFERENCES AND NOTES

1. B. F. Skinner, *Verbal Behavior* (Appleton-Century-Crofts, 1957).
2. N. Chomsky, A review of B. F. Skinner's verbal behavior. *Language* **35**, 26–58 (1959).
3. S. Pinker, *The Language Instinct: How the Mind Creates Language* (Morrow, 1994).
4. M. Tomasello, "The key is social cognition" in *Language in Mind: Advances in the Study of Language and Thought*, D. Gentner, S. Goldin-Meadow, Eds. (MIT Press, 2003), pp. 47–58.
5. A. E. Goldberg, *Constructions at Work: The Nature of Generalization in Language* (Oxford Univ. Press, 2006).
6. J. Elman *et al.*, *Rethinking Innateness: A Connectionist Perspective on Development* (MIT Press, 1996).
7. E. Thelen, L. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press, 1994).
8. R. N. Aslin, J. F. Werker, J. L. Morgan, Innate phonetic boundaries revisited. *J. Acoust. Soc. Am.* **112**, 1257–1260 (2002). doi: 10.1121/1.1501904; pmid: 12398431
9. E. S. Spelke, K. D. Kinzler, Core knowledge. *Dev. Sci.* **10**, 89–96 (2007). doi: 10.1111/j.1467-7687.2007.00569.x; pmid: 17181705
10. C. Rowland, *Understanding Child Language Acquisition* (Routledge, 2013). doi: 10.4324/9780203776025
11. A. Vaswani *et al.*, "Attention is all you need" in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Inc., 2017), pp. 6000–6010.
12. A. Radford *et al.*, Language models are unsupervised multitask learners. *OpenAI Blog* **1**, 9 (2019).
13. OpenAI *et al.*, GPT-4 technical report. arXiv:2303.08774 [cs.CL] (4 March 2024).
14. W. K. Vong, W. Wang, A. E. Orhan, B. M. Lake, Grounded language acquisition through the eyes and ears of a single child. *Science* **383**, 504–511 (2024). doi: 10.1126/science.adi1374; pmid: 38300999
15. J. Gilkerson *et al.*, Mapping the early language environment using all-day recordings and automated analysis. *Am. J. Speech Lang. Pathol.* **26**, 248–265 (2017). doi: 10.1044/2016_AJSLP-15-0169; pmid: 28418456
16. E. Bergelson *et al.*, Everyday language input and production in 1,001 children from six continents. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2300671120 (2023). doi: 10.1073/pnas.2300671120; pmid: 38085754
17. H. Raviv *et al.*, The First 1,000 Days (1kD) Project - collecting and analyzing an ultra-dense naturalistic dataset of human baby development. bioRxiv 2026.03.19.712982 [Preprint] (2026); https://doi.org/10.64898/2026.03.19.712982.
18. J. Gilkerson, J. A. Richards, "The LENA natural language study" (Technical Report LTR-02-2, LENA Foundation, 2008).
19. M. Casillas, A. Cristia, A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. *Collabra Psychol.* **5**, 24 (2019). doi: 10.1525/collabra.209
20. B. Long *et al.*, The BabyView dataset: High-resolution egocentric videos of infants' and young children's everyday experiences. arXiv:2406.10447 [cs.CV] (22 July 2025).
21. J. Sullivan, M. Mei, A. Perfors, E. Wojcik, M. C. Frank, SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind (Camb.)* **5**, 20–29 (2021). doi: 10.1162/opmi_a_00039; pmid: 34485795
22. B. MacWhinney, The CHILDES project: Tools for analyzing talk. *Child Lang. Teach. Ther.* **8**, 217–218 (1992). doi: 10.1177/026565909200800211
23. K. Adolph, Video as data: From transient behavior to tangible recording. *APS Obs.* **29**, 23–25 (2016). pmid: 28736493
24. D. Roy *et al.*, "The Human Speechome Project" in *Symbol Grounding and Beyond*, Lecture Notes in Computer Science, vol. 4211, P. Vogt, Y. Sugita, E. Tuci, C. Nehaniv, Eds. (Springer, 2006), pp. 192–196.
25. M. Huelin Gorriz, M. Takigawa, D. Bendor, The role of experience in prioritizing hippocampal replay. *Nat. Commun.* **14**, 8157 (2023). doi: 10.1038/s41467-023-43939-z; pmid: 38071221
26. A. K. Gillespie *et al.*, Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron* **109**, 3149–3163.e6 (2021). doi: 10.1016/j.neuron.2021.07.029; pmid: 34450026
27. P. K. Kuhl, Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843 (2004). doi: 10.1038/nrn1533; pmid: 15496861
28. P. D. Eimas, E. R. Siqueland, P. Jusczyk, J. Vigorito, Speech perception in infants. *Science* **171**, 303–306 (1971). doi: 10.1126/science.171.3968.303; pmid: 5538846
29. P. K. Kuhl, Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* **50**, 93–107 (1991). doi: 10.3758/BF03212211; pmid: 1945741
30. P. K. Kuhl *et al.*, Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* **9**, F13–F21 (2006). doi: 10.1111/j.1467-7687.2006.00468.x; pmid: 16472309
31. P. K. Kuhl, B. T. Conboy, D. Padden, T. Nelson, J. Pruitt, Early speech perception and later language development: Implications for the "Critical Period." *Lang. Learn. Dev.* **1**, 237–264 (2005). doi: 10.1207/s15473341lld0103&4_2

Downloaded from https://www.science.org at Princeton University on November 09, 2026

32. D. Swingley, R. N. Aslin, Spoken word recognition and lexical representation in very young children. *Cognition* **76**, 147–166 (2000). doi: 10.1016/S0010-0277(00)00081-0; pmid: 10856741
33. J. F. Werker, R. C. Tees, Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* **7**, 49–63 (1984). doi: 10.1016/S0163-6383(84)80022-3
34. D. R. Mandel, P. W. Jusczyk, D. B. Pisoni, Infants' recognition of the sound patterns of their own names. *Psychol. Sci.* **6**, 314–317 (1995). doi: 10.1111/j.1467-9280.1995.tb00517.x; pmid: 25152566
35. L. Fenson *et al.*, *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual* (Brookes Publishing Co., ed. 2, 2007).
36. W. Deng, S. Zhou, J. Shu, J. Wang, L. Wang, IndexTTS: An industrial-level controllable and efficient zero-shot text-to-speech system. arXiv:2502.05512 [cs.SD] (8 February 2025).
37. M. Lavechin *et al.*, Modeling early phonetic acquisition from child-centered audio data. *Cognition* **245**, 105734 (2024). doi: 10.1016/j.cognition.2024.105734; pmid: 38335906
38. M. Poli, T. Schatz, E. Dupoux, M. Lavechin, Modeling the initial state of early phonetic learning in infants. *Lang. Dev. Res.* **5**, 1–34 (2024).
39. N. H. Feldman, S. Goldwater, E. Dupoux, T. Schatz, Do infants really learn phonetic categories? *Open Mind* **5**, 113–131 (2021). doi: 10.1162/opmi_a_00046; pmid: 35024527
40. T. Schatz, N. H. Feldman, S. Goldwater, X.-N. Cao, E. Dupoux, Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2001844118 (2021). doi: 10.1073/pnas.2001844118; pmid: 33510040
41. A. Liu, H.-J. Chang, M. Auli, W.-N. Hsu, J. Glass, "DinoSR: self-distillation and online clustering for self-supervised speech representation learning" in *Proceedings of the 37th International Conference on Neural Information Processing Systems* (Neural Information Processing Systems Foundation, 2023), pp. 58346–58362.
42. E. Kharitonov *et al.*, "Text-free prosody-aware generative spoken language modeling" in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Association for Computational Linguistics, 2022), pp. 8666–8681.
43. U. Hasson, J. Chen, C. J. Honey, Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn. Sci.* **19**, 304–313 (2015). doi: 10.1016/j.tics.2015.04.006; pmid: 25980649
44. J. L. Elman, Finding structure in time. *Cogn. Sci.* **14**, 179–211 (1990). doi: 10.1207/s15516709cog1402_1
45. Z. Goriely, P. Buttery, "BabyLM's first words: word segmentation as a phonological probing task" in *Proceedings of the 29th Conference on Computational Natural Language Learning* (Association for Computational Linguistics, 2025), pp. 522–539. doi: 10.18653/v1/2025.conll-1.34
46. R. L. Gómez, R. R. Bootzin, L. Nadel, Naps promote abstraction in language-learning infants. *Psychol. Sci.* **17**, 670–674 (2006). doi: 10.1111/j.1467-9280.2006.01764.x; pmid: 16913948
47. B. McMurray, The acquisition of speech categories: Beyond perceptual narrowing, beyond unsupervised learning and beyond infancy. *Lang. Cogn. Neurosci.* **38**, 419–445 (2023). doi: 10.1080/23273798.2022.2105367; pmid: 38425732
48. O. Räsänen, H. Rasilo, A joint model of word segmentation and meaning acquisition through cross-situational learning. *Psychol. Rev.* **122**, 792–829 (2015). doi: 10.1037/a0039702; pmid: 26437151
49. J. F. Werker, H. H. Yeung, Infant speech perception bootstraps word learning. *Trends Cogn. Sci.* **9**, 519–527 (2005). doi: 10.1016/j.tics.2005.09.003; pmid: 16202639
50. M. C. Frank, M. Braginsky, D. Yurovsky, V. A. Marchman, Wordbank: An open repository for developmental vocabulary data. *J. Child Lang.* **44**, 677–694 (2017). doi: 10.1017/S0305000916000209; pmid: 27189114
51. B. C. Galland, B. J. Taylor, D. E. Elder, P. Herbison, Normal sleep patterns in infants and children: A systematic review of observational studies. *Sleep Med. Rev.* **16**, 213–222 (2012). doi: 10.1016/j.smrv.2011.06.001; pmid: 21784676
52. R. L. Gómez, J. O. Edgin, Sleep as a window into early neural development: Shifts in sleep-dependent learning effects across early childhood. *Child Dev. Perspect.* **9**, 183–189 (2015). doi: 10.1111/cdep.12130; pmid: 26557155
53. R. L. Gómez, K. Esterline, "A role for sleep in understanding language acquisition" in *International Handbook of Language Acquisition*, J. S. Horst, J. von Koss Torkildsen, Eds. (Routledge, 2019), pp. 128–150. doi: 10.4324/9781315110622-8
54. M. Friedrich, I. Wilhelm, J. Born, A. D. Friederici, Generalization of word meanings during infant sleep. *Nat. Commun.* **6**, 6004 (2015). doi: 10.1038/ncomms7004; pmid: 25633407
55. A. K. Lee, M. A. Wilson, Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002). doi: 10.1016/S0896-6273(02)01096-6; pmid: 12495631
56. Y. Liu, R. J. Dolan, Z. Kurth-Nelson, T. E. J. Behrens, Human replay spontaneously reorganizes experience. *Cell* **178**, 640–652.e14 (2019). doi: 10.1016/j.cell.2019.06.012; pmid: 31280961
57. I. Wilhelm, A. Prehn-Kristensen, J. Born, Sleep-dependent memory consolidation—What can be learnt from children? *Neurosci. Biobehav. Rev.* **36**, 1718–1728 (2012). doi: 10.1016/j.neubiorev.2012.03.002; pmid: 22430027
58. V. Krenz, A. Alink, B. Roozendaal, T. Sommer, L. Schwabe, Memory boost for recurring emotional events is driven by initial amygdala response promoting stable neocortical patterns across repetitions. *J. Neurosci.* **45**, e2406232025 (2025). doi: 10.1523/JNEUROSCI.2406-23.2025; pmid: 39947923
59. A. Ben-Yakov, V. Smith, R. Henson, The limited reach of surprise: Evidence against effects of surprise on memory for preceding elements of an event. *Psychon. Bull. Rev.* **29**, 1053–1064 (2022). doi: 10.3758/s13423-021-01954-5; pmid: 34173187
60. D. L. Schacter, The seven sins of memory. Insights from psychology and cognitive neuroscience. *Am. Psychol.* **54**, 182–203 (1999). doi: 10.1037/0003-066X.54.3.182; pmid: 10199218
61. E. Dupoux, Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition* **173**, 43–59 (2018). doi: 10.1016/j.cognition.2017.11.008; pmid: 29324240
62. C. F. Rowland *et al.*, Constructing language: A framework for explaining acquisition. *Trends Cogn. Sci.* **30**, 26–39 (2026). doi: 10.1016/j.tics.2025.05.015; pmid: 40562678
63. A. Weisleder, A. Fernald, Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* **24**, 2143–2152 (2013). doi: 10.1177/0956797613488145; pmid: 24022649
64. J. Y. Song, K. Demuth, J. Morgan, Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* **128**, 389–400 (2010). doi: 10.1121/1.3419786; pmid: 20649233
65. E. D. Thiessen, E. A. Hill, J. R. Saffran, Infant-directed speech facilitates word segmentation. *Infancy* **7**, 53–71 (2005). doi: 10.1207/s15327078in0701_5; pmid: 33430544
66. T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, G. Hinton, Backpropagation and the brain. *Nat. Rev. Neurosci.* **21**, 335–346 (2020). doi: 10.1038/s41583-020-0277-3; pmid: 32303713
67. A. Goldstein *et al.*, Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* **25**, 369–380 (2022). doi: 10.1038/s41593-022-01026-4; pmid: 35260860
68. J. E. Kosie, C. Lew-Williams, Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Dev. Sci.* **27**, e13515 (2024). doi: 10.1111/desc.13515; pmid: 38618899
69. E. Bates, J. C. Goodman, On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Lang. Cogn. Process.* **12**, 507–584 (1997). doi: 10.1080/016909697386628
70. S. Donnelly, E. Kidd, J. Verkuilen, C. Rowland, The separability of early vocabulary and grammar knowledge. *J. Mem. Lang.* **141**, 104586 (2025). doi: 10.1016/j.jml.2024.104586
71. U. Hasson, S. A. Nastase, A. Goldstein, Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron* **105**, 416–434 (2020). doi: 10.1016/j.neuron.2019.12.002; pmid: 32027833
72. L. B. Smith, E. Thelen, Development as a dynamic system. *Trends Cogn. Sci.* **7**, 343–348 (2003). doi: 10.1016/S1364-6613(03)00156-6; pmid: 12907229
73. H. Raviv, A. Tsyhanov, The First 1,000 Days: modeling language learning from individual children's everyday speech, Version 1.0. Zenodo (2026); https://doi.org/10.5281/zenodo.22649720.

ACKNOWLEDGMENTS

We thank the 1kD families for their extraordinary generosity and participation. We thank I. Kopaliani, J. Wiggins, and M. Ratliff from Princeton University for their exceptional institutional support. We also thank M. Faryna, S. Samko, and O. Chikhiro of Sigma Software and R. Beaty and D. Johnson of Beaty Consultancy for their contributions to the project's software infrastructure. We especially thank C. Greenberg from Princeton University for her work with participating families and B. Ryan for her work on child detection algorithms. We thank three anonymous reviewers for their constructive comments and suggestions, which helped improve the manuscript. We also thank M. C. Frank (Stanford University) for helpful discussions and feedback. During manuscript preparation, we used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and clarity improvements. We have reviewed and edited all the generated content and accept full responsibility for the manuscript. **Funding:** This work was supported by Wellcome Leap (H.R., C.L.-W., K.G., L.H., U.H.); the McGregor-Girand Charitable Endowment (C.L.-W.); Princeton Language and Intelligence (U.H., C.L.-W., A.T.); Princeton Precision Health (H.R.); Data-Driven Social Science, Princeton University (U.H., C.L.-W.); AI Lab, Princeton University (U.H., C.L.-W., A.T.); and CV Starr Fellowship, Princeton Neuroscience Institute (H.R.). The content of this publication and any related materials are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies. **Author contributions:** Conceptualization: H.R., U.H., L.H., C.L.-W.; Data curation: O.R., B.C., H.R., A.A., U.H., L.H., C.L.-W., K.G.; Formal analysis: H.W., H.R., U.H., A.T.; Funding acquisition: U.H., L.H., C.L.-W.; Investigation: B.C., H.R., A.A., L.H., C.L.-W., K.G.; Methodology: O.R., B.C., H.W., H.R., A.A., U.H., L.H., C.L.-W., T.R., K.G., A.T.; Project administration: L.H.; Software: O.R., B.C., H.W., H.R., T.R., A.T.; Supervision: U.H., L.H., C.L.-W.; Validation: H.R., U.H., A.T.; Visualization: H.R., U.H.; Writing – original draft: H.R., U.H.; Writing – review & editing: O.R., L.H., C.L.-W., T.R., A.T. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** The 1kD data are not publicly available because of the private nature of long-duration audio and video

Downloaded from https://www.science.org at Princeton University on November 09, 2026

recordings collected inside participating families' homes and the potential sensitivity of associated derived materials. Access is granted through a controlled research process for approved uses that fall within the scope of the original consent and institutional review board–approved protocol. Requests are reviewed for scientific scope, requested data layers, export limits, and dissemination risk. Approved researchers work within the secure 1kD research environment under a data-use agreement and project-specific access conditions. Code supporting the learning-model framework, analysis pipeline, and figure generation is available at https://github.com/hassonlab/1kd_learning_model and Zenodo (*73*). No new materials were generated in this study. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aeh9311
Materials and Methods; Supplementary Text; Figs. S1 to S11; References (*74*–*83*); MDAR Reproducibility Checklist

Submitted 11 May 2026; accepted 11 September 2026

10.1126/science.aeh9311

Downloaded from https://www.science.org at Princeton University on November 09, 2026

Science
AAAS

The first 1000 days: Modeling language learning from individual children's everyday speech

Hadas Raviv, Arkadii Tsyhanov, Kira Gousios, Aja Altenhof, Haocheng Wang, Berlin Chen, Ofri Raviv, Tal Rosenwein, Casey Lew-Williams, Liat Hasenfratz, and Uri Hasson

Science **394** (6824), eaeh9311. DOI: 10.1126/science.aeh9311

Editor's summary

Large language models are normally trained on massive collections of text before they can produce verbal output, but then they can create sophisticated pieces of writing. By contrast, human infants learn from much more limited verbal input and then produce a few words at a time before ramping up. To simulate human children's experience, Raviv *et al*. trained a machine learning model on dense longitudinal recordings of words heard by individual human infants, with daytime input equivalent to what a single child heard in one day followed by overnight replay of the cumulative input up to that point (see the Perspective by Harmon and Rowland). With this training approach, each child's model learned at about the same rate as the human child, supporting the real-world relevance of this modeling approach. —Yevgeniya Nusinovich

View the article online
https://www.science.org/doi/10.1126/science.aeh9311
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at Princeton University on November 09, 2026

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works