The infant's view redefines the problem of referential uncertainty in early word learning

Chen Yu[a,b], Yayun Zhang[a], Lauren K. Slone[c], and Linda B. Smith[a,d,1]

[a]Department of Psychology, The University of Texas at Austin, Austin, TX 78712; [b]Department of Psychological and Brain Sciences, Cognitive Science Program, Indiana University Bloomington, Bloomington, IN 47405; [c]Psychology Department, Hope College, Holland, MI 49423; and [d]School of Psychology, University of East Anglia, NR4 7TJ Norwich, Norfolk, United Kingdom

Contributed by Linda B. Smith; received October 27, 2021; accepted November 1, 2021; reviewed by Jenny Saffran and John Trueswell

The learning of first object names is deemed a hard problem due to the uncertainty inherent in mapping a heard name to the intended referent in a cluttered and variable world. However, human infants readily solve this problem. Despite considerable theoretical discussion, relatively little is known about the uncertainty infants face in the real world. We used head-mounted eye tracking during parent–infant toy play and quantified the uncertainty by measuring the distribution of infant attention to the potential referents when a parent named both familiar and unfamiliar toy objects. The results show that infant gaze upon hearing an object name is often directed to a single referent which is equally likely to be a wrong competitor or the intended target. This bimodal gaze distribution clarifies and redefines the uncertainty problem and constrains possible solutions.

language | statistical learning | development | infancy

Before their first birthday, infants recognize a few object names, looking reliably to the referent upon hearing the name (1, 2). By their second birthday, infants recognize and produce several hundreds of object names (3–5). Infants learn those early object names by linking heard words to seen objects (6, 7). This is a computationally hard problem because at the moment of hearing a name, there are likely many potential referents in the vicinity of the young learner (8, 9). This puzzle of referential uncertainty, first posed by Quine (10), has defined theoretical debates and empirical research on early word learning for the past 50 y (5, 9, 11). Despite considerable research effort, there is no accepted solution nor explanation of how infants rapidly learn object names.

Competing accounts begin with different assumptions. By one view (e.g., ref. 6), referential uncertainty in the input is rampant and intractable without strong internal perceptual, cognitive, and linguistic constraints on nameable categories that reduce uncertainty. Some constraints, such as linking novel names to whole objects but not their parts, may be innate (12). But many others, such as biases to link novel names to novel objects, are known not to be fully in play at the start of learning but instead emerge as a product of early learning experiences (13–16). By the second view, certainty rather than uncertainty is common because parents use social cues (points, gaze, and gestures, etc.) to indicate the referent (17–20). Young learners could discard ambiguous naming moments without the support of social cues and instead focus on the socially guided transparent moments. By the third view, the learning environment contains a mix of more and less ambiguous naming events (21–24), and infants learn name–object correspondences from taking all the data from such a mix, either through cross-situational statistical learning (22, 25, 26) or by hypothesis testing (21). Each of these alternative views has been supported by compelling results derived from laboratory experiments. However, those experiments are designed under different uncertainty assumptions. The relevance and likelihood that those assumptions match the uncertainty in everyday infant experience are under question (27).

The field needs a way to directly quantify the uncertainty in learning environments. Although previous research (28, 29) has attempted to directly measure uncertainty, proper and accurate measurement is not straightforward. Because referential uncertainty depends on the number of potential referents in the learning environment, an intuitive approach is to quantify the number of potential objects in the vicinity of the infant when a name is heard. Some studies analyzed third-person-view videos of parent–child interactions and measured uncertainty in terms of adult observers' ability to guess the parent-intended referent from the visual information alone (28, 29). One problem with this approach is that all the objects available in the vicinity are not necessarily the objects in the infant's field of view (30–32). Some studies have used adult-judgments but egocentric video thus measuring ambiguity from the infant point of view (31, 33). However, the psychology of adults guessing the parent-intended referent even from the infant point of view may not be the same as that of infants.

Our starting premise is that referential uncertainty is a property of the learner and that it needs to be quantified with respect to the information that is selected and attended to by the learner. If a young learner, even in a messy environment, typically directs gaze to one or a very few things upon hearing a name, then the uncertainty for the learner would be considerably less than the total number of available referents in the environment. This selectivity could be the result of intrinsic constraints, visual saliency of a potential target, parent social cues, or the structure of the visual environment. If, however,

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Significance

Human infants learn object names by mapping heard names to perceived things. The key challenge to this learning is referential uncertainty: For any heard name, there are many candidate referents with variable perceptual properties. Previous experimental research on how infants solve this problem has been based on different assumptions about the nature of this uncertainty. The present study quantifies the degrees of referential uncertainty by measuring infants' real-time attention to the potential referents when parents name unfamiliar objects. We discovered an unexpected bimodal distribution of infant attention to the intended referents in parent speech during everyday parent–infant interactions, which redefines the century-long puzzle of referential uncertainty as a problem in the sampling of data for learning.

Author contributions: C.Y. and L.B.S. designed research; C.Y., Y.Z., and L.K.S. performed research; C.Y., Y.Z., L.K.S., and L.B.S. analyzed data; and C.Y. and L.B.S. wrote the paper.

Reviewers: J.S., University of Wisconsin–Madison; and J.T., University of Pennsylvania.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: smith4@indiana.edu.

Published December 21, 2021.

the novice learner, upon hearing an object name, directs gaze randomly to many possible referents, then the uncertainty within and across naming events would be high (34). The computational problem that infant learners need to solve depends on the number of possible referents from the learner's point of view. Accordingly, breaking current barriers to explanation depends on determining the exact properties of the uncertainty with respect to the potential referents that enter the infant's learning system for internal computations. Such a determination would provide insight into the mechanisms underlying early object name learning and would provide guidance for empirical studies designed to compare potential learning mechanisms (35). Toward these goals, we quantified referential uncertainty from the learner's perspective using infant gaze upon hearing an object name as the measure of the possible referents being considered by the infant's learning system.

Results

Because our goal was to quantify the uncertainty from the infant's point of view without prejudice as to extant hypotheses, we arranged a context with many nameable objects, 24 toys (Fig. 1*A*) of likely interest to infants. The toys varied along many dimensions, just as do the objects that infants encounter at home, including multiple instances of the same and similar categories that varied in their likely familiarity. We recruited 36 infants (M = 19.3, SD = 2.8) and their parents and asked parents to interact with their infant and whatever toys they wished. Parents were not told to name the objects nor that the study was about language. Infants were free to move in the space (Fig. 1*B*). Each dyad played with the set of toys for over 6 min (M = 6.32 min, SD = 2.58 min), and together the whole group provided a corpus containing 224 min of gaze data in a free-flowing interaction. There were 1,508 spontaneous parent naming events in the corpus. The specific names provided by parents for the individual objects were chosen by and varied across participants.

Prior research (36) indicates that the proportion of time that infant gaze is directed to the parent-intended referent after hearing its name predicts learning of the name–object mapping (Fig. 1*C*). Therefore, to quantify the uncertainty of individual naming events, we used the proportion of time that infant attention was directed to the intended referent. We operationally defined a temporal window beginning at the start of a naming utterance and measured the mean proportion of time during that specified window that infant gaze was directed to the intended referent. Because the observed proportions necessarily vary with the selected window size, we conducted the main analyses using a 3-s window motivated by prior work (37, 38) but also report the results for varied window sizes.

During a measured window, infant gaze could be directed to a single object or distributed across multiple potential referents. If there were strong social, perceptual, or cognitive constraints on referent selection, infants could attend correctly and consistently to the parent-intended referent. Alternatively, infants could either shift gaze among the available referents considering a variety of possible and therefore mostly wrong referents for each naming event, or they could wrongly focus on a single referent. Fig. 2 shows four illustrative outcome distributions

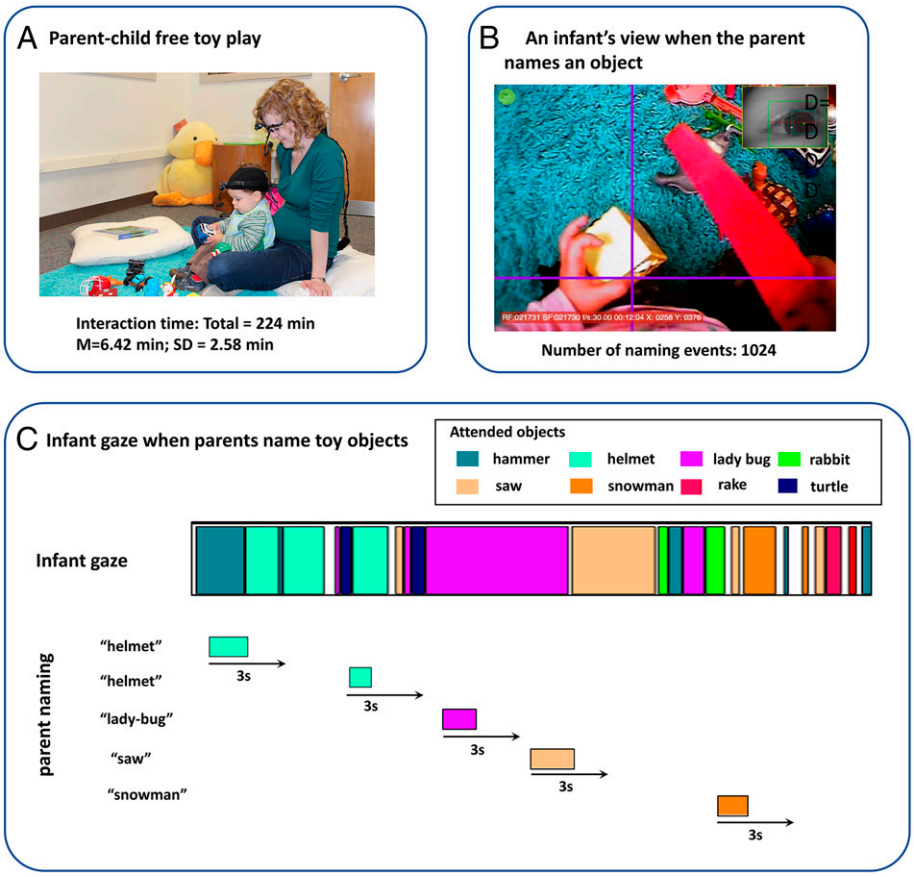

Fig. 1. Experimental setup and data. (*A*) Infants and their parents played freely with a set of toys. Infants wore a head-mounted eye tracker which recorded gaze data from the infant's perspective. (*B*) Infant gaze, indicated by a crosshair in the infant's first-person view, was on an object when hearing a to-be-learned name. (*C*) Gaze directed to different objects is illustrated by different colors in the infant gaze stream. Infant gaze was temporally aligned with parent naming. A set window, 3 s in the main analyses, from the onset of each naming instance was used.

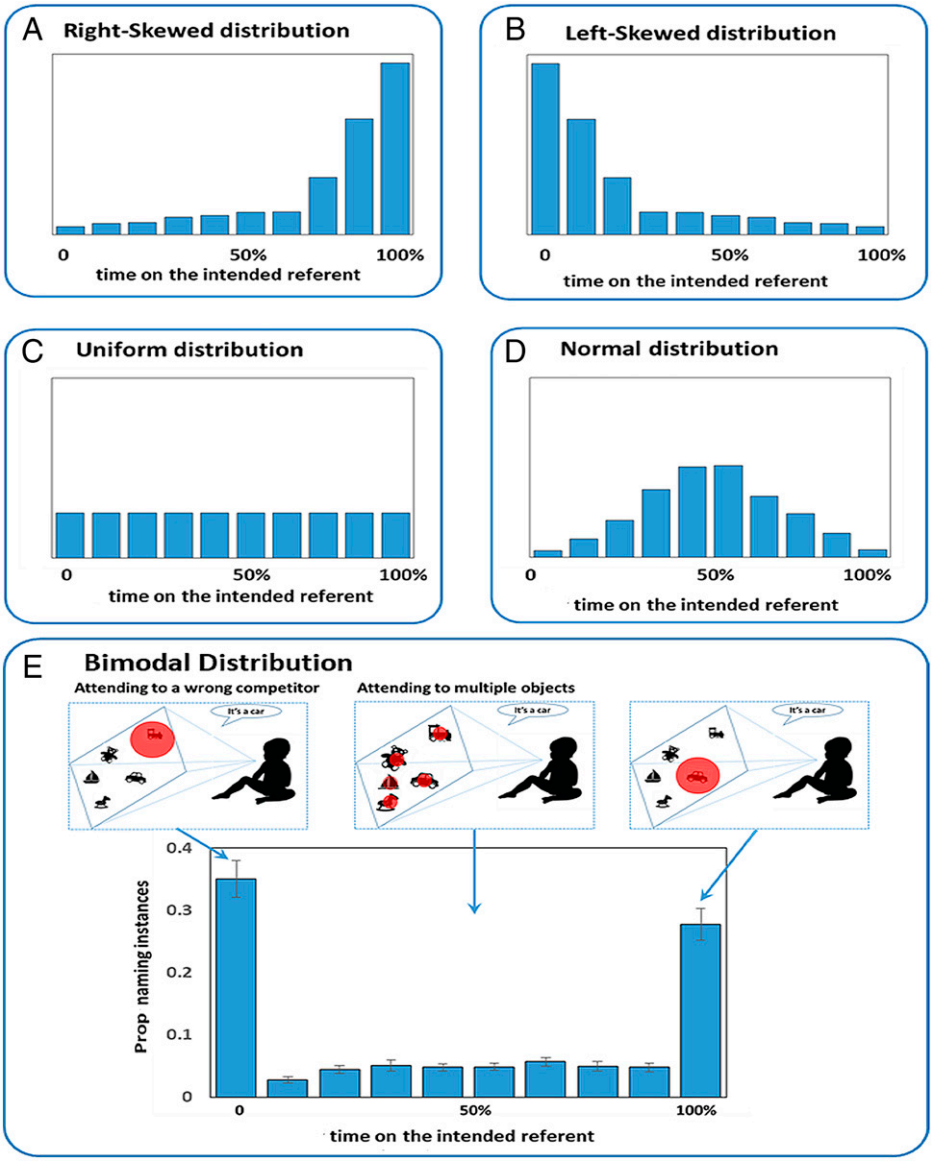

Fig. 2. Distributions of the naming events in joint play that characterize referential uncertainty in the learning environment. (*A*–*D*) Four hypothesized theoretical distributions. (*A*) A skewed distribution in which most of the naming events contain a high degree of uncertainty. (*B*) A skewed distribution in which most of the naming events are transparent with a low degree of uncertainty. (*C*) A uniform distribution in which the naming events with different degrees of uncertainty occur equally. (*D*) A normal distribution in which most of the naming events are at the middle level containing some degree of uncertainty, and fewer instances are at either side containing extremely high or low uncertainty. (*E*) The observed bimodal distribution of naming instances observed in the present data. Some naming events (bin 0%) have high uncertainty because infants attended to a wrong competitor in the entire time during and after hearing a name; other naming events have low uncertainty (bin 100%) because infants gazed at the target object the whole time when hearing its name; and the rest of naming instances fall in between the two extreme cases (between bin 0 and 100%) as infants attended to multiple potential referents including not only the target object but also other competitors. The distribution shows the M and SD for each bin determined across all 1,024 naming events.

based on different assumptions about referential uncertainty in the literature. Fig. 2*A* illustrates a high-certainty distribution: If the environment, parents' behaviors, and/or infants' internal biases and past experiences conspire to yield transparent naming events, then the likelihood that gaze is directed to the intended referent upon hearing an object name should be high for most of the naming events. Fig. 2*B* shows the opposite extreme of high uncertainty: Infant gaze is primarily directed to wrong referents. This pattern would occur if a learner had no knowledge and randomly selected objects upon hearing a name. By statistical learning accounts, learning should be slow but possible under this degree of uncertainty (39, 40). These two hypothesized distributions are illustrated as following a Zipfian distribution, which has been consistently shown to characterize human behavior (41) including the distribution of word frequencies in infant-directed speech (26, 42–45) and the duration of infant looks to a single object (46, 47). Fig. 2 *C* and *D* show possible distributions at midlevels of uncertainty where there is a mix of degrees of uncertainty among the naming events, some of which are associated with high uncertainty and others with low uncertainty. These intermediate levels of uncertainty have been commonly instantiated in experimental studies and conform to some proposals about the statistical learning of word–referent correspondences (23, 48–50).

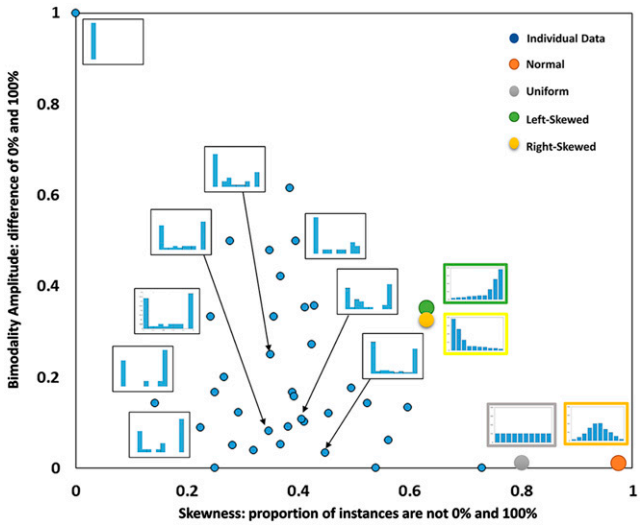

Fig. 3. Distributions of naming events extracted from individual infants and four hypothesized distributions. The *x* axis is defined to measure certainty (either 0 or 100%) and the *y* axis is defined to measure whether 0 and 100% instances are roughly equal. Despite expected individual differences in free play, distributions from individual infants share certain properties defined based on bimodal amplitude (*y* axis) and skewness (*x* axis) that are not characterized by the four hypothesized distributions.

The observed distribution (Fig. 2*E*) differed from the illustrated possibilities shown in Fig. 2. During a 3-s window after the onset of a naming event, infant gaze shows a pattern consistent with a combination of high certainty (Fig. 2*A*) and high uncertainty (Fig. 2*B*) as to the intended referent. Infants either looked to the intended referent during and after a naming event or, equally often, looked to wrong objects. On 82% of the naming events that fell into bin 0% (no gaze directed to the intended referent), infants primarily directed gaze to a single object, spending twice as much time on a single wrong object than on the second-most attended object (M = 63.5%, SD = 10.2%). The two extremes—looking 100% in the 3-s window to the intended referent or 100% to competitors (and most often to a single competitor)—accounted for 65% of all parent naming events. For the remaining 35% of the naming events, infants distributed gaze to multiple objects including at least some visual attention to the intended referent. The bimodal distribution implies that the factors that constrain infant referent selection upon hearing a name are all or none: Either the infant is 100% certain of the correct referent and looks to it or the infant does not direct gaze to the correct referent at all.

The Bimodal Distribution at the Level of Individual Dyads. The bimodal distribution shown in Fig. 2*E* has two unique and defining properties: 1) The two modes together cover a significant percentage of all naming events (M = 65.35%, SD = 14.2%); and 2) the two modes are roughly equal—gaze to the right referent or to a wrong one. To determine how well this overall distribution fits individual parent–infant dyads, we calculated "bimodality amplitude," a measure of difference between the two modes, and "skewness," the percentage of naming instances that are not in the two modes (neither 0 nor 100% to the intended referent). Small values in bimodal amplitude mean similar likelihood between naming instances in which infant gaze is on the correct or incorrect object. Small values of skewness mean that most naming events in the bimodal distribution are in the two modes (0 or 100%). Fig. 3 shows the bimodal amplitude and skewness of the distributions derived from individual dyads as well as those from the four hypothetical distributions (Fig. 2 *A*–*D*). Given the context of spontaneous naming by parents and free-flowing interactions, one might expect considerable individual differences. However (Fig. 3), there are marked commonalities among the distributions for individual dyads. Most dyads fall in a cluster centered around the lower-left area of the plot, because most naming instances within these individual dyads consisted of instances of infant gaze directed entirely to the intended referent or never to the intended referent. Only one dyad created a relatively large number (greater than 60%) of naming events that were neither 0 nor 100%. Some dyads created more 100% events and others more 0% events, but the data points from most dyads are close to the *x* axis, showing that most of them created similar numbers of 0 and 100% events (with one exception at the upper-right corner in which all naming instances are 0%). We also varied the duration of the temporal window after a naming onset to be 500, 1,000, 2,000, 3,000, 4,000, and 5,000 ms. The infants' attention distribution consistently shows a bimodal distribution within the shorter or longer durations after naming onset (Fig. 4*A*). Thus, at both the group and individual levels, the bimodal pattern appears to be a stable property of young learners' responses to parent naming.

The Bimodal Distribution as a Function of Parent Naming in the Task. These distributions could arise as the product of repeated naming experiences during play. Parents might frequently name a single object, enabling the infant to learn that name during play (and thus direct gaze correctly to the object). Thus, the two modes could reflect the differences between looking behavior to more and less frequently named objects and to looking behavior during the first and second halves of the play session. If so, the overall bimodal distributions accumulated from all the naming events in a play session might reflect a combination of an early right-skewed distribution (Fig. 2*B*) and a later left-skewed distribution (Fig. 2*A*).

Each dyad chose different toys for play at each moment and spent different amounts of time on the chosen toys. Parents generated, on average, 41.61 naming events in each play session and the frequency of naming varied considerably among the dyads (SD = 25.23), with one parent generating 110 naming events but another just 4 naming events. Within a dyad, different objects were named more frequently than others. For example, the most frequently named object was mentioned 7.72 times in a play session, and on average there were two available objects (different for different parents) never named in parent speech. Fig. 4*B* shows the frequency distribution of naming events aggregated over the rank order of objects named within individual dyads. Two analyses of infants looking to the intended referent were conducted to examine whether naming within the task impacts the bimodal distribution, one focused on the effects of the frequency of individual object names and one focused on changes in looking behavior from earlier to later in the interaction.

To examine overall frequency effects, we divided naming events into two groups: 1) a high-frequency group containing the naming instances from the top six most frequently named objects for each dyad—different toys for different dyads but the top six across all dyads account for all 65% of all naming events; and 2) a low-frequency group containing the rest of the naming events for the remaining 18 toys that were each infrequently named by the parent. As shown in Fig. 4*C*, there is no difference between the two groups and the two distributions are similar to the overall distribution shown in Fig. 2*E* (Mann–Whitney *U* test, $U = 29$, $P > 0.30$). Thus, the distribution of looks to the intended referent do not vary systematically with the frequency of naming experiences within the play session.

To examine potential changes in looking behavior as a function of time in the interaction and thus as a function ot the

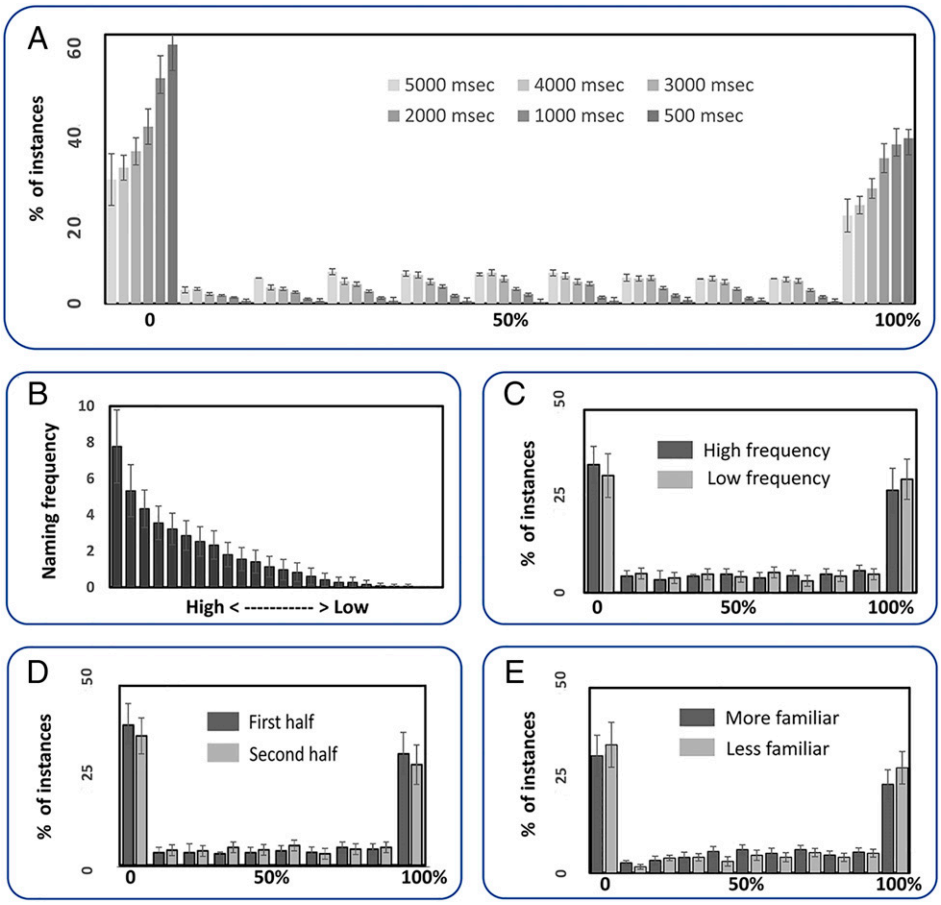

Fig. 4. Additional analyses relevant to determining factors potentially responsible for the observed bimodal distribution. (*A*) The bimodal distributions with different temporal windows, ranging from 500 to 5,000 ms with M and SD calculated over the 1,024 naming events. (*B*) The skewed distribution of object-naming frequency. Naming of individual objects is sorted based on the frequency for the individual dyad (that is, the most frequently named object for one dyad is aggregated with the most frequent for another, even though these are not the same toy). Error bars indicate the SD of the frequency distribution among the dyads. (*C*) The distributions of infant attention when hearing high frequency and low frequency names. Distribution, M, SD determined over 1,024 naming events. (*D*) The distributions of infant attention when hearing high-frequency and low-frequency names. Distribution, M, and SD are determined over the 1,024 naming events. (*D*) The distributions of infant attention when hearing high frequency and low frequency names. Distribution, M, SD determined over 1,024 naming events. (*E*) The bimodal distributions on more and less familiar objects. Distribution, M, and SD are determined over the 1,024 naming events.

number of in-task naming experiences, we focused on the parent naming of the top six named objects for each dyad, the objects that were named sufficiently to show a potential first-half to second-half difference in looking behavior. The top six objects within a dyad were, on average, named 27.08 times in a play session. Given that each name in the top list was mentioned multiple times in a play session, infants could respond differently when they first heard a specific name but change their looking behavior after hearing the same name repeatedly with increasing familiarity to that name. To examine whether infant attention to the most frequently named objects changed over the course of a play session, we divided the naming instances of the top six named objects within each dyad into two groups based on whether each naming instance occurred in the first half or second half of the play session. We then calculated the distributions of infant attention for all naming events for these six objects occurring in the first and second halves of the play session. The two histograms in Fig. 4*D* show no difference of infant attention for the high frequency (more familiar) and less frequent (less familiar) object names in the play session (Mann–Whitney U test, $U = 35$, $P > 0.30$). This result is also inconsistent with a proposal that parents first name novel objects in a play context in a way that makes them more transparent, supporting rapid learning that can withstand subsequent ambiguous name–object correspondences (51, 52).

The Bimodal Distribution as a Function of Name and Object Familiarity. The toy set used included some highly typical and some unusual items. Some names, categories, and individual objects were likely more familiar to the infants than others prior to the experiment and thus this preexperiment familiarity may influence infant attention. Accordingly, we estimated the likelihood of infant familiarity with heard object names as follows: We first identified all the names in parent speech that referred to toy objects. In cases in which parents used more than one name to refer to the same object (e.g., the Rubik's cube was referred to as "Rubik's," "block," and "cube"), each name was evaluated separately. We then used the MacArthur-Bates Communicative Development Inventories report (3) provided by parents before the experiment to determine whether each of the parent-used labels was likely known or unknown for their child. At the subject level, parents reported that their infants knew 38% of the names mentioned in parent speech (SD = 9%, range 0 to 89%). At the item level, the most known object name (car) was reported to be known to the infant by 75% of parents, and the least known parent-uttered name

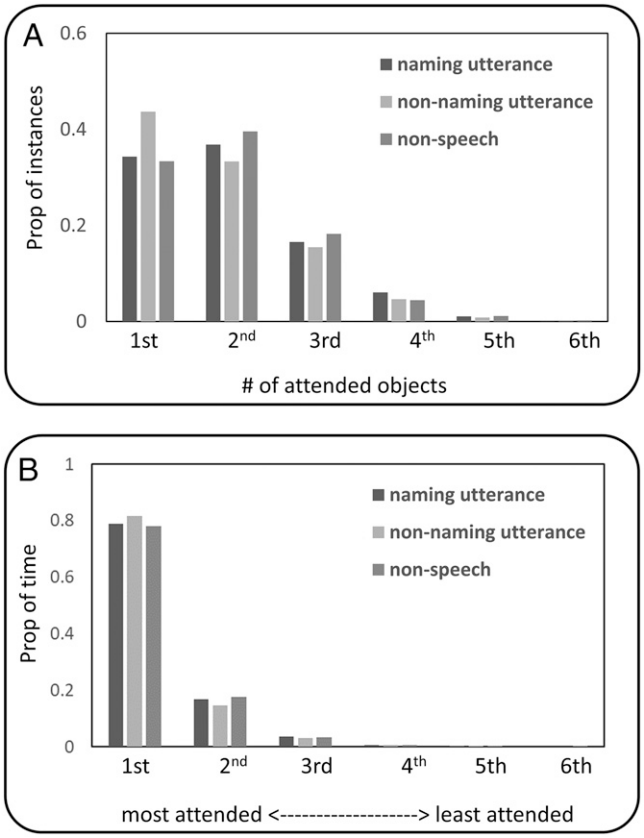

Fig. 5. Comparison of three types of moments in parent–infant play: 1) naming utterance: a temporal window defined from the onset of a naming utterance; 2) nonnaming utterance: a temporal window defined from the onset of a spoken utterance without any toy name; and 3) nonspeech: a temporal window defined from the offset of a spoken utterance. (*A*) The number of attended objects within a 3-s window. (*B*) The distribution of attention from the most attended to the least attended object within a 3-s window.

(mantis) was reported to be known by 0% of the infants by their parents. Given the wide variability among dyads in toys named, we divided data from individual infants by parent report of infant knowledge of that name. As shown in Fig. 4*E*, there is no difference between the two groups; the bimodal distribution is obtained for both likely more familiar and less familiar object names (Mann–Whitney U test, $U = 35$, $P > 0.30$).

Our approach to the role of partial knowledge in referential uncertainty has its limitations; neither parent-report measure is likely a perfect measure. Further, learning about heard names and about the kinds of things to which they refer is a continuous and accumulated process: An object or a word would not suddenly become known from unknown. Nonetheless, the familiarity of a particular name and a particular toy appears to have little impact on how infants distribute gaze upon hearing an object name. The main results are these: The bimodal distributional properties of infant attention to the named objects are consistently observed, not varying with the frequency of heard names in the task, with when individual naming events occurred, nor with the estimated degree of infant familiarity with the object names used by parents.

Naming versus Nonnaming Moments. So far, we showed that infants select and attend to a single object within a short time window after hearing a label with the typically single selected object being the intended referent or some other object. Is this attentional behavior to a single object a response specific to labeling moments? Fig. 5 compares naming utterances with utterances containing nonlabeling speech and nonspeech moments. In all the three situations, infants attended to only a very few objects (e.g., Fig. 5*A*), and spent a majority of time on a single object (Fig. 5*B*). The findings suggest that throughout the play session, infants show similar patterns of fixating a single object within a short time window (3 s, etc.) and thus that this extended attention to a single object is not specific to labeling moments. This looking pattern of infants could be an intrinsic property of their attentional system or reflect the properties of attentional cues in the world which may include parent behavior and talk other than labeling as well as interesting events.

Discussion

Referential uncertainty is the property of a learner in an environment. Taking this perspective, the present results suggest that the field would benefit from a radical rethinking of the construct of referential ambiguity and the psychology that makes infants rapid and robust learners of object names. The specific contributions of this study are the findings that 1) after each naming instance, infants tend to direct gaze predominantly to just one potential referent; 2) across naming instances, infants sometimes select the correct referent and sometimes a wrong distractor, selecting the correct target over one-third of the time, which is much better than would be expected if they randomly selected any object in their view; and 3) this bimodal pattern of looking to the correct referent or to a wrong competitor across naming instances is ubiquitous across individual infants, repetitions of naming, and potential partial knowledge. In this discussion, we present a conceptualization of referential uncertainty, its implications for understanding learning environments, the underlying learning mechanisms, and individual differences in vocabulary development.

Referential Uncertainty Is about How Learners Sample Data from the World. Referential uncertainty has typically been conceptualized as a fact about the external world that presents a problem to be solved by the learner (8, 10). However, at each moment, all the information available in the external world is not available to the infant because the infant selectively samples that information and that sampling depends on the infant's own momentary internal state (27, 53). The present results show that this sampling process leads to "signature" distributional properties of sampled objects for a heard name that differ from previous assumptions (44, 54). Gaze can be directed to only one spatial location at a time; visual sampling thus may be best construed as a "decision" at the level of looking behavior itself, driven by the interactions among top-down processes, external saliences, and recent sensorimotor and social events. All the proposed factors that have been shown to be relevant in observational and experimental studies—parent naming, parent gestures, clutter, linguistic and contextual cues, the infant's current state, and potentially more—may all be relevant and may compete in complex ways. What we have shown here is that these competing forces are most often resolved in an all-or-none "winner-take-all" solution to the direction of the gaze. The observed bimodal distribution suggests that for infants, about one-third of the time the suite of relevant factors conspire to lead the infant to the intended referent and about one-third of the time they do not. The fact that none of the examined individual factors (frequency of naming or familiarity) were shown to influence naming may not mean that they are not relevant. Rather, the processes that control visual sampling in the moment for infants may be complex, nonlinear, and involve many co-occurring signals. Notice that if this

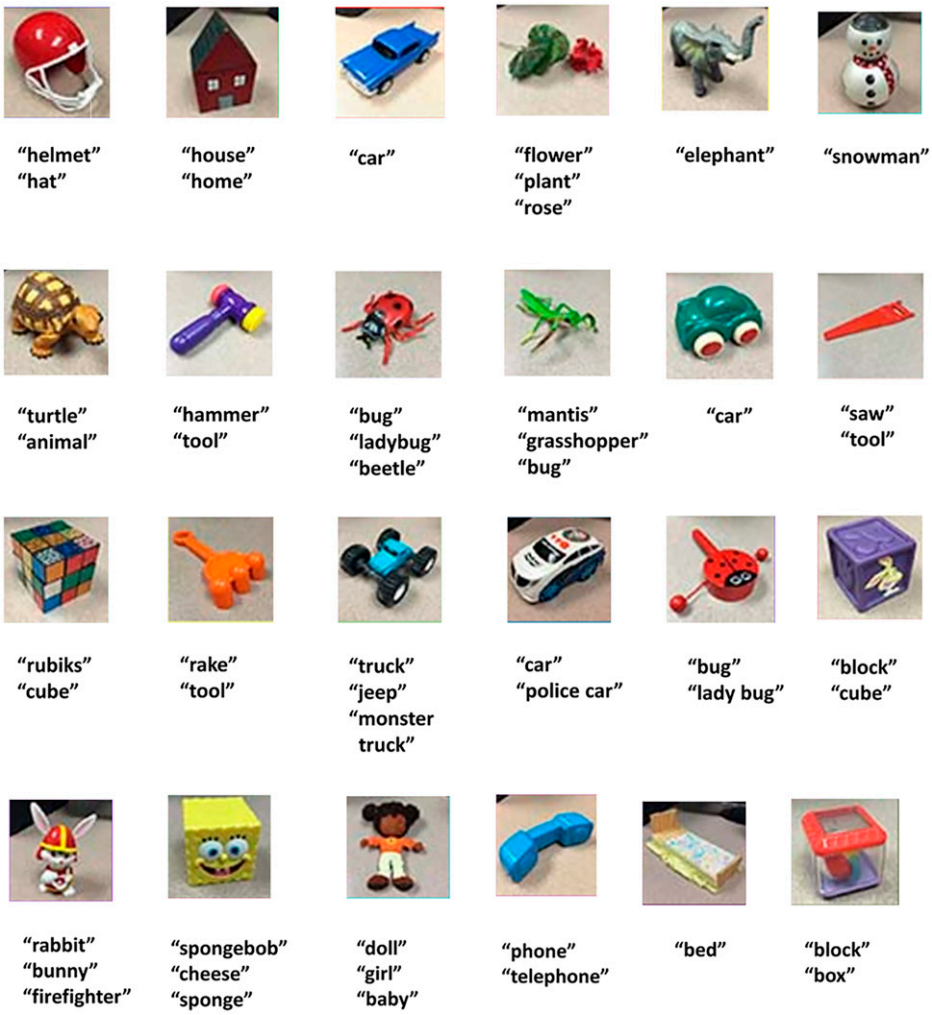

Fig. 6. Twenty-four toy objects and corresponding names used by parents.

conceptualization is right, referential uncertainty experienced by individual infants may be a driving factor of individual differences in lexical learning.

How the Sampled Data with a Bimodal Distribution May Be Processed for Word Learning. As noted by many theorists of word learning, the world presents many potential referents for any heard word (8, 10, 11, 39). However, the relevance of this uncertainty in the external world to infants' word learning is filtered through the psychology of the young learner. Extensive data—observations from the real world and experiments in the laboratory—show that infants, in general, are robust learners of object names (23). To do so, the learning mechanisms that infants use must find an efficient way to operate on the sampled data for early word learning.

In the present context of toy play, a context often considered optimal for object name learning, the data for learning a word through repeated naming events are a mix of correct and misleading co-occurrences. What kind of learning mechanism learns well from this data structure? An optimal mechanism may be one that does not commit too quickly to a word–referent pairing given potential spurious associations but also one that settles fast enough to yield efficient learning. Framed in this way, infant learning of object names may be conceptualized as an exploitation–exploration problem (55, 56). A learner who exploits previous learning and persists in sampling the same information may not find the optimal or even correct solution; a learner who randomly samples new information may discover new information but never settle on a stable or correct solution. Considerable research across many different domains has suggested that one optimal way to sample data is to use a hybrid approach—staying with what seemed to work before but meanwhile being open-minded to explore something new. A bimodal distribution may give rise to an exploitation–exploration balance that sufficiently explores alternative referents and in so doing efficiently finds the optimal—that is, correct—referent.

This exploration–exploitation approach may also align with internal memory processes. Recent evidence indicates that conflicting data create internal competitions which in turn strengthen associations (57, 58). Through active inhibition of the many different individual spurious associations, a learning system can build and consolidate robust memories for the correct name–object pairings (48, 57). At the computational level, competitive processes can be implemented in all three of the major computational accounts of early object name learning: associative learning, statistical learning, and hypothesis testing. One benefit of the evident exploration–exploitation sampling pattern—not committing too soon to a referent early on, keeping the system open to learning—is to elicit competitive processes that strengthen the association of the statistically more prevalent and intended referent to the name.

How the Bimodal Distribution Is Created. The problem of early object–name learning has been formalized as follows: Given that an unknown (or not completely known) object name is heard, how does the infant resolve the in-the-moment ambiguity? Our discussions of sampling and possible mechanisms focused not on the information presented by the world in a single trial but on the information sampled from the infant's perspective. In principle, at each moment, sampling could be a response of the infant to an instance of parent naming and thus occur only as a consequence of a labeling event. Prior research has shown that infant attention when hearing a label is affected by a range of variables, from prior interest in the object category (59) to the visual salience of competitor objects (60) to overall novelty (61). Alternatively, the observed sampling pattern may reflect a general property of infant visual attention as influenced by several factors and thus not a specific response to a heard name. Infants are in the process of learning much more than object names, including learning about the functions and affordances of objects. During toy play, infant look durations to a target are a mix of short looks (46, 47, 62), with long looks often but not always accompanied by manual actions (47, 62). The long looks, often referred to as instances of sustained attention, are also strongly associated with parent speech (62, 63) and with infant learning (36). Thus, long looks during toy play may sometimes emerge from factors other than naming but elicit parent naming as well as sometimes being a response to a heard name. A key question for future research is how the coupled behaviors of infants and parents interact with respect to the observed bimodal distribution.

The findings reported here are observed in toy play, a free-flowing and common everyday context. Two critical questions are whether the same observations would be obtained in other contexts (e.g., mealtime or book reading), and how much they vary across different contexts and across different individuals (64). Further, in addition to naming the toys, parents also talked about those objects and described their properties; infant sampling of visual information is relevant to and likely influenced by those parent behaviors as well. Our conjecture is that the balance of exploration and exploitation observed in the present results is a general property of infant attentional decisions and looking behavior, in part because that behavior is driven by a complex system of interacting factors and because a balance in staying and shifting attention may be optimal for naïve learners with much to learn in many different domains (56). All this suggests the value of studying and linking the statistical properties of infant visual sampling in naturalistic contexts to learning in different contexts and domains.

Conclusion

Referential uncertainty has been conceptualized as a property of the environment. We proposed that uncertainty be conceptualized as a momentary property of the learner in an environment. Using this framework, we found a ubiquitous pattern of near-equal certainty and uncertainty—looking only to the intended referent or looking to a single nonreferent. These are the data on which contemporary proposed mechanisms of hypothesis testing and statistical learning must operate. Further, laboratory studies should not be designed and evaluated based on presumed statistics but on the statistics from the infant's perspective in complex contexts that are like everyday experiences. More radically, to resolve the disagreement and to unify different views of early word learning, the field may benefit by putting aside the construct of referential uncertainty and reframing the research task as one of understanding how young infants sample information from the environment, how that sampling process interacts with possible learning mechanisms, how sampling varies across individuals, and how sampling processes change with development.

Materials and Methods

Experimental Setup. The 36 dyads (parents and infants) were recruited from an outreach event and the sample was broadly representative of Monroe County, Indiana (84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Other) and consisted of predominantly working- and middle-class families. Recruitment, consenting procedures, and the research protocol were approved by the Human Subjects and Institutional Review Board at Indiana University (protocol no. 0808000094). Parents gave consent for the infant's participation and remained with the infant throughout their participation. The laboratory environment was set up like home with everyday furniture, including a couch, chairs, pillows, a TV stand, lamps, an eating area. and a play area. For this study, the play area had 24 toys randomly spread on the floor. Parents sat on the carpet, a posture they reported to be natural and comfortable. Infants began the session sitting on the carpet next to their parent. However, the infants were free to move and, for instance, might move to sit on their parent's lap, or crawl, or stand up and walk around for a short period of time before coming back to resume joint toy play with their parents. Parents were told to play—and allow their child to play—with the toys as they usually would. There were no additional constraints on parent behavior or instructions to parents about what they should do, for example, which toys they should select, what actions they should generate, or what they should say to their child. The goal was to create a family-friendly environment for free-flowing toy play like that typical in the home. Both participants wore head-mounted eye trackers from Positive Science LLC. Each eye-tracking system includes an infrared camera—mounted on the head and pointed to the right eye of the participant that records eye images, and a scene camera capturing the first-person view from the participant's perspective. The scene camera's visual field is 90°, providing a broad view but one less than the full visual field (~170°). Each eye-tracking system recorded both the egocentric-view video and gaze direction (x and y) in that view, with a sampling rate of 30 Hz. Three additional high-resolution cameras (recording rate 30 frames per second) were mounted to surround the play area and provided third-person environmental views that were independent of participants' movements. Parent speech was recorded from a microphone mounted next to one of the environmental cameras.

Materials. Twenty-four everyday toys were selected. Based on normative data, their names were expected not to be in the vocabulary of the infants but to be known to parents. Parents were free, of course, to name the objects by any name, and they sometimes used more than one name to refer to the same toy. The toy objects included, for example, SpongeBob, football helmet, monster truck, bed, Rubik's cube, police car, turtle, rabbit, ladybug, elephant, mantis, saw, and shovel. A list of toys and object names used in parent speech is shown in Fig. 6. Additional toys were used to engage the child during the placement of the eye tracker and its calibration.

Procedure. The procedure for placing the eye tracker on the infant closely follows methods in a previously reported study (65). One experimenter played with the infant while a second placed the eye-tracking gear low on the forehead of the infant at a moment when the infant was engaged with a toy used only for this phase of the experiment. The third experimenter controlled the computer to ensure data recording. The first experimenter directed the infant's attention toward a toy used only for calibration while the second experimenter recorded the attended moment that was used in later eye-tracking calibration. This procedure was repeated 15 times with the calibration toy placed in various locations in the play area. Parents were told that the goal of the experiments was to study how parents and infants played with objects during free play. Therefore, they were asked to engage their infants with the toys and to do so as naturally as possible. They were not told that we were interested in naming events, nor were they instructed to name the objects. They played for up to 10 min, ending earlier if the infant became fussy or dislocated the eye-tracking device.

Data and Data Processing. The head-mounted eye trackers recorded gaze data at a rate of 30 frames per second and only gaze data from infants were used in the present study. There were about 224 min of interaction, yielding potentially 403,200 gaze data points. Not all participants provided eye-tracking data for the entire session. Roughly 25% of frames from infants were not codable with respect to regions of interest (ROIs, defined in the next

paragraph); this was due to 10% eye-tracking failure (e.g., occlusion of pupil image in the eye camera of the infant eye tracker) and the rest was due to the infant's being off-task (looking elsewhere than defined ROIs). In total, the method uses microbehavioral analyses with over 9,712 gaze data points from each infant. We annotated gaze and speech data during toy play, from which we derived measures to examine both quantity and quality of parent naming.

Gaze data. The 24 ROIs were the 24 toy objects in play at a time. ROIs were coded manually frame by frame, determining whether the crosshair indicating gaze direction overlapped any portion of an object and, if so, on which object. Each child provided a gaze data stream as shown in Fig. 1*C*. A second coder independently coded a randomly selected 10% of the frames with 95% agreement.

Naming events. Parent speech was transcribed into spoken utterances, among which those containing the names of the toys were designated as naming events. Naming events were defined as the whole utterance in which a name was embedded and were on average 1.5 s in duration. Each naming event was coded as a triplet <onset, offset, name>. An example data stream of naming events is shown in Fig. 1*C*.

Data Availability. The coded data analyzed for this paper have been deposited in the OSF (https://osf.io/mvz78/).

Raw video data are not publicly available because they contain information that could compromise the privacy of research participants (e.g., young children and their families). Questions may be directed to the corresponding author or to C.Y.

ACKNOWLEDGMENTS. We thank Steven Elmlinger, Seth Foster, Daniel Pearcy, and Charlene Tay for collection of the data; Seth Foster and Tian (Linger) Xu for developing data management and processing software; and Tian Xu, Drew Abney, Catalina Suarez, Sara Schroer, Julia Yurkovic-Harding, and Sven Bambach for fruitful discussions. This work was funded by NIH Grants R01HD074601 and R01HD093792, NSF Grant BCS-1842817, and Indiana University's Emerging Area of Research Initiative.

1. E. Bergelson, D. Swingley, At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3253–3258 (2012).
2. R. Tincoff, P. W. Jusczyk, Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* **10**, 172–175 (1999).
3. L. Fenson *et al.*, *MacArthur-Bates Communicative Development Inventories* (Brookes Publishing, Baltimore, MD, ed. 2, 2005).
4. L. Bloom, E. Tinker, C. Margulis, The words children learn: Evidence against a noun bias in early vocabularies. *Cogn. Dev.* **8**, 431–450 (1993).
5. R. M. Golinkoff, C. B. Mervis, K. Hirsh-Pasek, Early object labels: The case for a developmental lexical principles framework. *J. Child Lang.* **21**, 125–155 (1994).
6. P. Bloom, *How Children Learn the Meaning of Words* (MIT Press, Cambridge, MA, 2000).
7. J. Macnamara, *Names for Things: A Study of Human Learning* (MIT Press, Cambridge, MA, 1982).
8. T. N. Medina, J. Snedeker, J. C. Trueswell, L. R. Gleitman, How words can and cannot be learned by observation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9014–9019 (2011).
9. L. Gleitman, K. Cassidy, R. Nappa, A. Papafragou, J. Trueswell, Hard words. *Lang. Learn. Dev.* **1**, 23–64 (2005).
10. W. Quine, *Word and Object* (MIT Press, Cambridge, MA, 1960).
11. L. B. Smith, S. H. Suanda, C. Yu, The unrealized promise of infant statistical word-referent learning. *Trends Cogn. Sci.* **18**, 251–258 (2014).
12. E. Markman, Constraints children place on word meanings. *Cogn. Sci.* **14**, 57–77 (1990).
13. J. Halberda, Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognit. Psychol.* **53**, 310–344 (2006).
14. E. M. Markman, G. F. Wachtel, Children's use of mutual exclusivity to constrain the meanings of words. *Cognit. Psychol.* **20**, 121–157 (1988).
15. C. Houston-Price, Z. Caloghiris, E. Raviglione, Language experience shapes the development of the mutual exclusivity bias. *Infancy* **15**, 125–150 (2010).
16. R. A. Bion, A. Borovsky, A. Fernald, Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition* **126**, 39–53 (2013).
17. D. A. Baldwin *et al.*, Infants' reliance on a social criterion for establishing word-object relations. *Child Dev.* **67**, 3135–3153 (1996).
18. N. Akhtar, M. Tomasello, "The social nature of words and word learning" in *Becoming a Word Learner: A Debate on Lexical Acquisition*, R. M. Golinkoff *et al.*, Eds. (Oxford University Press, 2000), pp. 115–135.
19. M. Carpenter, K. Nagell, M. Tomasello, G. Butterworth, C. Moore, Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* **63**, i–vi, 1–143 (1998).
20. R. W. Cheung, C. Hartley, P. Monaghan, Caregivers use gesture contingently to support word learning. *Dev. Sci.* **24**, e13098 (2021).
21. J. C. Trueswell, T. N. Medina, A. Hafri, L. R. Gleitman, Propose but verify: Fast mapping meets cross-situational word learning. *Cognit. Psychol.* **66**, 126–156 (2013).
22. C. Yu, L. B. Smith, Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* **18**, 414–420 (2007).
23. L. Smith, C. Yu, Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* **106**, 1558–1568 (2008).
24. J. S. Stevens, L. R. Gleitman, J. C. Trueswell, C. Yang, The pursuit of word meanings. *Cogn. Sci.* **41**, 638–676 (2017).
25. K. Smith, A. Smith, R. Blythe, P. Vogt, "Cross-situational learning: A mathematical approach" in *Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, P. Vogt, Y. Sugita, E. Tuci, C. Nehaniv, Eds. (Springer, Berlin, 2006), pp. 31–44.
26. A. T. Hendrickson, A. Perfors, Cross-situational learning in a Zipfian environment. *Cognition* **189**, 11–22 (2019).
27. L. B. Smith, S. Jayaraman, E. Clerkin, C. Yu, The developing infant creates a curriculum for statistical learning. *Trends Cogn. Sci.* **22**, 325–336 (2018).
28. E. A. Cartmill *et al.*, Quality of early parent input predicts child vocabulary 3 years later. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11278–11283 (2013).
29. J. C. Trueswell *et al.*, Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition* **148**, 117–135 (2016).
30. C. Yu, L. B. Smith, Embodied attention and word learning by toddlers. *Cognition* **125**, 244–262 (2012).
31. D. Yurovsky, L. B. Smith, C. Yu, Statistical word learning at scale: The baby's view is better. *Dev. Sci.* **16**, 959–966 (2013).
32. L. Smith, C. Yu, H. Yoshida, C. M. Fausey, Contributions of head-mounted cameras to studying the visual environments of infants and young children. *J. Cogn. Dev.* **16**, 407–419 (2015).
33. Y. Zhang, D. Yurovsky, C. Yu, Cross-situational learning from ambiguous egocentric input is a continuous process: Evidence using the human simulation paradigm. *Cogn. Sci. (Hauppauge)* **45**, e13010 (2021).
34. L. B. Smith, S. H. Suanda, C. Yu, The unrealized promise of infant statistical word-referent learning. *Trends Cogn. Sci.* **18**, 251–258 (2014).
35. A. Kingstone, D. Smilek, J. D. Eastwood, Cognitive ethology: A new approach for studying human cognition. *Br. J. Psychol.* **99**, 317–340 (2008).
36. C. Yu, S. H. Suanda, L. B. Smith, Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Dev. Sci.* **22**, e12735 (2019).
37. A. F. Pereira, L. B. Smith, C. Yu, A bottom-up view of toddler word learning. *Psychon. Bull. Rev.* **21**, 178–185 (2014).
38. H. A. Ruff, K. R. Lawson, Development of sustained, focused attention in young children during free play. *Dev. Psychol.* **26**, 85–93 (1990).
39. R. A. Blythe, A. D. M. Smith, K. Smith, Word learning under infinite uncertainty. *Cognition* **151**, 18–27 (2016).
40. B. McMurray, T. P. Ellis, K. S. Apfelbaum, How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear Hear.* **40**, 961–980 (2019).
41. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Ravenio Books, 2016).
42. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
43. D. Swingley, C. Humphrey, Quantitative linguistic predictors of infants' learning of specific English words. *Child Dev.* **89**, 1247–1267 (2018).
44. C. Kurumada, S. C. Meylan, M. C. Frank, Zipfian frequency distributions facilitate word segmentation in context. *Cognition* **127**, 439–453 (2013).
45. P. Vogt, Exploring the robustness of cross-situational learning under Zipfian distributions. *Cogn. Sci.* **36**, 726–739 (2012).
46. J. I. Borjon, D. H. Abney, C. Yu, L. B. Smith, Head and eyes: Looking behavior in 12- to 24-month-old infants. *J. Vis.* **21**, 18 (2021).
47. C. Yu, L. B. Smith, The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* **26**, 1235–1240 (2016).
48. C. Yu, L. B. Smith, Modeling cross-situational word-referent learning: Prior questions. *Psychol. Rev.* **119**, 21–39 (2012).
49. H. A. Vlach, S. P. Johnson, Memory constraints on infants' cross-situational statistical learning. *Cognition* **127**, 375–382 (2013).
50. R. M. Scott, C. Fisher, 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition* **122**, 163–180 (2012).
51. S. C. Berens, J. S. Horst, C. M. Bird, Cross-situational learning is supported by propose-but-verify hypothesis testing. *Curr. Biol.* **28**, 1132–1136.e5 (2018).
52. K. Woodard, L. R. Gleitman, J. C. Trueswell, Two- and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Lang. Learn. Dev.* **12**, 252–261 (2016).
53. M. Zettersten, J. R. Saffran, Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Dev. Sci.* **24**, e13064 (2019).
54. S. Bambach, D. Crandall, L. Smith, C. Yu, "Toddler-inspired visual object learning" in *Advances in Neural Information Processing Systems*, S. Bengio *et al.*, Eds. (Neural Information Processing Systems Foundation, 2018), pp. 1209–1218.
55. R. C. Wilson, E. Bonawitz, V. D. Costa, R. B. Ebitz, Balancing exploration and exploitation with information and randomization. *Curr. Opin. Behav. Sci.* **38**, 49–56 (2021).

56. A. Gopnik, Childhood as a solution to explore-exploit tensions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190502 (2020).
57. D. Yurovsky, C. Yu, L. B. Smith, Competitive processes in cross-situational word learning. *Cogn. Sci.* **37**, 891–921 (2013).
58. A. A. Bhat, J. P. Spencer, L. K. Samuelson, Word-object learning via visual exploration in space (WOLVES): A neural process model of cross-situational word learning. *Psychol. Rev.*, 10.1037/rev0000313 (2021).
59. L. Ackermann, R. Hepach, N. Mani, Children learn words easier when they are interested in the category to which the word belongs. *Dev. Sci.* **23**, e12915 (2020).
60. R. Pomper, J. R. Saffran, Familiar object salience affects novel word learning. *Child Dev.* **90**, e246–e262 (2019).
61. J. S. Horst, L. K. Samuelson, S. C. Kucker, B. McMurray, What's new? Children prefer novelty in referent selection. *Cognition* **118**, 234–244 (2011).
62. C. Suarez-Rivera, L. B. Smith, C. Yu, Multimodal parent behaviors within joint attention support sustained attention in infants. *Dev. Psychol.* **55**, 96–109 (2019).
63. S. E. Schroer, C. Yu, "Multimodal attention creates the visual input for infant word learning" (IEEE International Conference of Development and Learning, 2021). https://doi.org/10.1109/ICDL49984.2021.9515676.
64. L. R. Masek *et al.*, Where language meets attention: How contingent interactions promote learning. *Dev. Rev.* **60**, 100961 (2021).
65. C. Yu, L. B. Smith, Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One* **8**, e79659 (2013).