Developmental changes in the speed of social attention in early word learning

Daniel Yurovsky[1], Anna Wade[2], Allison M Kraus[3], Grace W. Gengoux[4], Antonio Y. Hardan[4], and Michael C. Frank[3]

[1]Department of Psychology, University of Chicago

[2]School of Medicine, University of California, San Francisco

[3]Department of Psychology, Stanford University

[4]School of Medicine, Stanford University

Author Note

Please address correspondence to:

Daniel Yurovsky
Department of Psychology
University of Chicago
5848 S University Avenue
Chicago IL, 60637

Email: yurovsky@uchicago.edu

Abstract

How do children learn words so rapidly? A powerful source of information about a new word's meaning is the set of social cues provided by its speaker (e.g. eye-gaze). Studies of children's use of social cues have tended to focus on the emergence of this ability in early infancy. We show, however, that this early-emerging ability has a long developmental trajectory: Slow, continuous improvements in speed of social information processing occur over the course of the first five years of life. This developing ability to allocate social attention is a significant bottleneck on early word learning—continuous changes in social information processing predict continuous changes in children's ability to learn new words. Further, we show that this bottleneck generalizes to children diagnosed with autism spectrum disorder, whose social information processing is atypical. These results describe a route by which increases in social expertise can lead to changes in language learning ability, and more generally highlight the dependence of developmental outcomes not on just the existence of particular competencies, but on their proficient use in complex contexts.

Developmental changes in the speed of social attention in early word learning

Children's first years are a time of rapid change. One striking development is children's growing mastery of their native language: The typical child will go from saying her first word shortly before she turns one to producing 4000–5000 words by age five (1). The pace and breadth of this transformation have led to a search for early-available, precocious mechanisms that support children's language acquisition. Perhaps because of the fundamentally discrete nature of the units of language itself, much of this search has focused specifically on pinpointing the earliest emergence of these mechanisms (2–4). However, learning outside the laboratory is controlled not by early availability of these mechanisms, but by their proficient use in complex natural environments.

Here we take as a case study children's use of social information to infer the meanings of new words. Though even the earliest vocabularies contain words belonging to many grammatical categories, concrete nouns make up a large proportion (5). While acquiring a fully adult-like meaning for any of these nouns likely unfolds over multiple encounters, the very first problem a child faces when hearing a new word is referential uncertainty: Does the word refer to something in the current situation, and if so, what (6–8)? A powerful source of information for resolving this uncertainty is available in the social cues provided by the speaker; knowing where a speaker is looking is helpful for knowing what they are communicating about. Consequently, a large body of research documents and explores young infants' ability to track a speaker's social cues and use them to infer the target of her reference (e.g., 9–12).

Much of the work in this research program has focused on discovering the earliest point of infants' competence in using social information. Underlying this focus is an implicit assumption that once the general ability to use social cues is demonstrated, children's proficiency with processing these cues is relatively high (e.g. 13–15). This research strategy stands in contrast to work in domains like visual and motor development, or even spoken word recognition, in which researchers have sought to measure continuous improvement in children's abilities as they develop (16–20). In these domains, continuous, quantitative changes can have as much impact on an infants' interaction with the world as qualitative changes (21). For example, the transition from crawling to walking leads to some advantages in children's ability to explore. However, infants' initial mobility after this transition is nowhere near as great as what they will achieve with another few months of practice (22). The emergence of the behavior is only the beginning.

Our goal in this article is to follow this same developmental strategy—of measuring continuous improvement—to understand how children use social information to learn new words. Using social information to resolve referential uncertainty is a highly time-sensitive process of continuous re-allocation of attention between the speaker and the objects in the context. Indeed, recent work has shown that rapid gaze-following occurs relatively rarely in natural parent-child interactions, and is difficult even for older children and adults under some circumstances (23–25). Thus, we hypothesize that competence in gaze following is not enough: the referential uncertainty problem should remain a problem in proportion to a children's developing ability to control their attention, process auditory and visual information, and hold this information in memory (26–28).

In three experiments, we test this hypothesis, showing that the developing ability to rapidly direct attention in social interactions is a bottleneck on children's word learning. We constructed videos that used novel words in a series of naturalistic object-focused dialogues and monologues. These videos were designed to be sufficiently difficult in their structure that in-the-moment disambiguation would pose a significant challenge to young learners, yet sufficiently simple that the repeated co-occurrence of novel words and the objects they label could allow for successful word learning. We measured children's eye-movements during viewing as an index of their online inferences about the current conversational referent, and then tested their retention of the words they learned via a series of forced-choice test trials. These rich time-course data allowed us to test two predictions: (1) although young infants show measurable competence in social information processing, this ability has a long developmental trajectory, and (2) this developing ability is linked to language learning: children who successfully attend to speakers' social cues are more likely to learn and retain the novel words they hear the speakers produce.

In Experiment 1, we measured developmental changes in social information processing and used these to predict word learning in typically developing children. In Experiment 2, we tested a sample of children diagnosed with autism spectrum disorder (ASD) in the same paradigm as Experiment 1. We showed that social attention was a bottleneck on their learning from these videos in the same way as for the typically developing children. In Experiment 3, we manipulated the timing of social information directly to test the causal role of fast social attention, again in typically developing children. In all three studies, we recruited children across a broad age range, giving us the power to see continuous changes in both social attention and novel word learning.

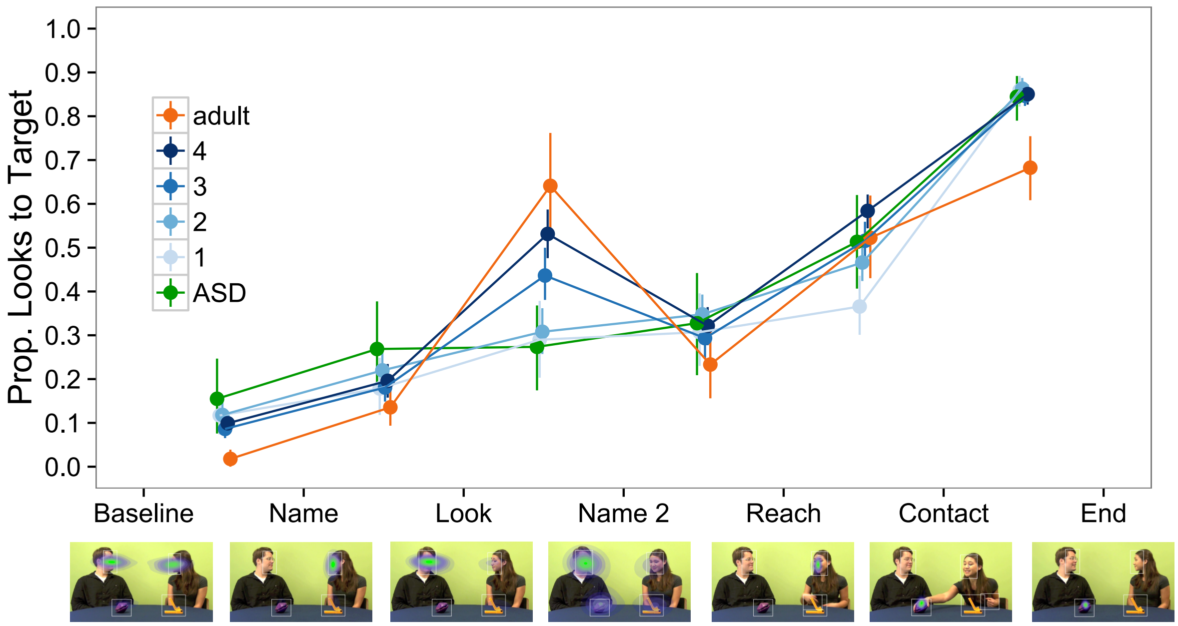

Figure 1. (top) Looking to Target vs. Competitor objects during the learning portion of Experiments 1 (typically-developing children and adults) and 2 (children with ASD), plotted by phase of the naming event. Points show means and error bars show 95% confidence interval across participants; points are offset on the horizontal to avoid overplotting. Colors indicate age in years for typically developing children. (bottom) Example frames from the first word learning dialogue in Experiment 1. Each image shows the regions of interest used for later analysis (white boxes) and a heat map of the distribution of all participants' points of gaze over time (brighter colors indicate more fixation; scale is constant across frames). See Supporting Information for a video representation.

Experiment 1: Social Attention and Word Learning

Experiment 1 was designed to estimate the developmental trajectory of children's online social information processing in complex interactions, and to ask whether this processing has downstream consequences for word learning. The experiment consisted of two sections: Learning and Test. During the learning section, participants watched a series of dialogues and monologues in which actors introduced and discussed two novel objects and two familiar objects. Each video contained two naming sequences (pictured in Fig. 1, bottom); these sequences were broken into a series of phases during which the speaker named, looked at, named again, and reached for a particular object, allowing for the separate measurement of name-, gaze-, and reaching-related changes of attention.

We first measured the proportion of time that participants spent looking at the target referent during the naming sequences (Fig. 1, top). We analyzed these gaze trajectories by fitting a mixed-effects model predicting looking at the target from naming phase, age, and their interaction. Children increased their looking to the target object over the course of learning trials, with above-baseline looks to the target in all phases after the second naming. For all age groups, looking to the target toy reached nearly 100% by the end of these events—after the speaker had made contact with the toy ($\beta_{name2-reach} = .18$, $t = 3.73$, $p < .001$; $\beta_{reach-contact} = .13$, $t = 2.56$, $p < .01$; $\beta_{contact-end} = .75$, $t = 15.8$, $p < .001$). The effect of age was not significant, but there was a significant interaction between age and phase—older children looked significantly more in the phases after the speaker's initial look and initial reach ($\beta_{look-name2} = .11$, $t = 7.38$, $p < .001$; $\beta_{reach-contact} = .08$, $t = 5.722$, $p < .001$). While the youngest children were only occasionally able to follow the speaker's social cues, older children and adults were much more consistent. Thus, the ability to process social information quickly and reliably in a naturalistic conversation improves markedly over the course of the first five years, and even further into adulthood.

After watching these naming events, children's learning of the novel words was tested using the looking-while-listening procedure (20, 29). On Test trials, children saw two toys

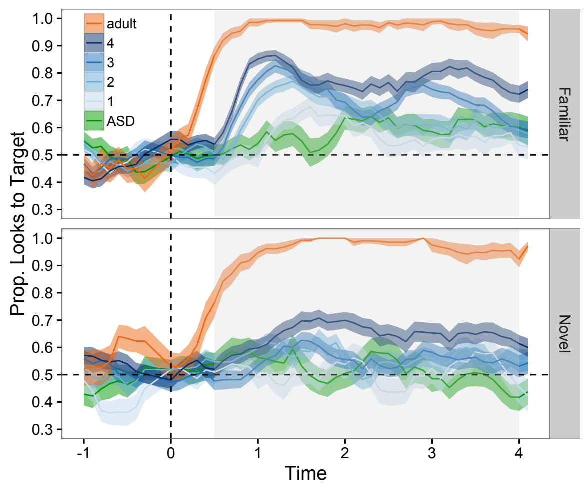

Figure 2. Children's and adults' looks to the Target and Competitor objects over the course of Test trials for typically developing children and adults (Experiment 1) and children with ASD (Experiment 2). Lines show age-group means, and shaded regions show standard errors. The light gray rectangle shows the window over which looking proportions were computed for subsequent statistical analyses.

on the screen and heard a voice asking them to find the target toy. On some trials, the target was a one of the Novel toys from the Learning trials (e.g. "fep"). On other trials, the target was a familiar object whose label is typically in the comprehension vocabularies of young children (e.g. "dog"). These Familiar trials allowed us to measure children's language comprehension more generally (Fig. 2). Children in all age groups successfully looked at Familiar referents at above chance levels (smallest $\mu_{1-year} = .57$, $t(28) = 4.05$, $p < .001$), and all but the one-year-olds reliably learned the Novel words (smallest $\mu_{2-year} = .56$, $t(51) = 3.96$, $p < .001$). A linear mixed-effects model showed that children performed better on Test trials over development ($\beta_{age} = .07$, $t = 7.89$, $p < .001$), and that the effect of age was greater for Familiar than Novel trials ($\beta_{age*novel} = -.03$, $t = -2.47$, $p < .05$, Fig. 2)

Thus, older children learned more from the same naming events. These children also more quickly followed the speakers' social gaze and reaches, and more quickly processed

Familiar words. All of these factors were independently, significantly correlated with learning ($r_{age}(196) = .29$, $p < .001$); $r_{familiar}(196) = .20$, $p < .01$; $r_{look-name2}(189) = .34$, $p < .001$; $r_{reach-contact}(191) = .17$, $p < .05$). Which factor was most responsible for improvements in learning? To answer this question, we fit a linear regression, predicting learning from age, familiar word processing, and looking to the target in the two relevant windows—after the look, and after the reach. Only looking to the target referent following the initial look reached significance ($\beta_{look-name2} = .14$, $t = 3.26$, $p < .01$), although age was marginal ($\beta_{age} = .02$, $z = 1.60$, $p = .11$). Because these predictors were all correlated, we also fit this same model after first residualizing out the effect of age on learning. In this model, gaze-following still remained highly significant ($\beta_{look-name2} = .11$, $t = 2.95$, $p < .01$).

Children who were unable to follow the speaker's social cues to determine the toy that was the topic of conversation were unable to learn that it was associated with the novel word in the discourse. Thus, age-related improvements in social attention, whatever their cause, play a powerful role in determining how much children learn from naming events throughout early childhood. The early emergence of this important competence in following social cues is no guarantee of its proficient use in complex, naturalistic contexts.

Experiment 2: Social Attention in Children with Autism

The naming events in Experiment 1 contained information sufficient to infer the meanings of the novel words at two timescales. First, children could learn the meanings within the interactions by following the informative social cues. Second, children could learn these meanings across the interactions by using co-occurrence statistics between the words and objects (30). In our data, typically developing children's social information processing predicted considerable variance in their word learning, suggesting that they may have learned largely via the social cues. Do children with atypical trajectories of social development rely more on co-occurrence statistics instead?

Children on the autism spectrum are one population whose strategy might be expected to differ. Deficits in social information processing in children with autism spectrum disorders have profound consequences for language learning (31, 32). One possibility is that because of these deficits, if children with autism learn in our paradigm, it would be due to cross-situational statistical learning. Another possibility however is that their social information processing impairments lie on a continuum with other, less extreme changes in social information processing that are still impactful for language learning (14). On this second account, their social information processing should be related to their learning outcomes, just as in typically-developing children.

To address this question, we tested a group of 40 2–8-year-old children diagnosed with autism spectrum disorder (ASD). These children did indeed process social information less efficiently—following the speaker's gaze no better than the youngest children in our sample (Fig. 1). They also learned the Novel words, and attended to the referents of Familiar words at approximately the same levels (Fig. 2). However, while age was uncorrelated with Novel word learning for the children with ASD ($r(28) = .11$, $p = .57$), following the speaker's social gaze was highly correlated with Novel word learning ($r(21) = .58$, $p < .01$), just as in the typically developing sample (Fig. 3). We again fit a linear mixed-effects regression, predicting Novel word learning from age, Familiar word processing, and looking to the Target in the two relevant windows—after the look, and after the reach. This model showed significant effects of Familiar word processing ($\beta_{familiar} = .51$, $t = 2.48$, $p < .05$) and gaze following ($\beta_{look-name2} = .25$, $t = 3.06$, $p < .01$).

Thus, social information processing is a bottleneck on word learning for children with autism spectrum disorder, just as it is for typically developing children. Although as a group these children's looking behavior during learning trials was different from the looking behavior of their typically developing peers, children with ASD who successfully learned the Novel words did so by following social cues.

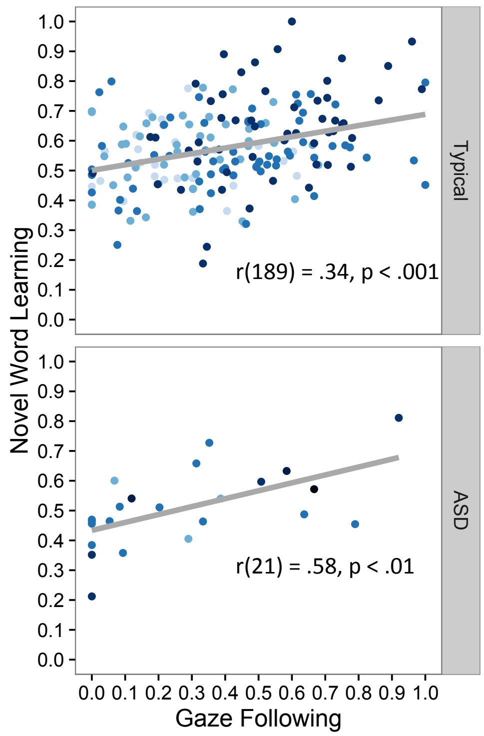

Figure 3. Correlations between gaze-following on Learning trials and accuracy on Novel Test trials for typically developing children in Experiment 1 and children with autism spectrum disorder (ASD) in Experiment 2. Darker colored points indicate older children.

Experiment 3: Varying Demands on Social Attention

The naming events in Experiments 1 and 2 provided children with a number of informative cues to the referents of the novel words: social gaze, speaker's manual interaction, and also cross-situational co-occurrence statistics. Our analyses showed that fast processing of social gaze was a strong predictor of children's ultimate word learning. In Experiment 3, we tested this prediction directly by manipulating the accessibility of the social cue.

In Experiment 3, the two Novel words were introduced in two different kinds of naming events. In Extended Hold events, the speaker made contact with and manipulated the Target toy for the duration of the naming event, providing an extended cue to the target of her referential intention. In contrast, in Brief Look events, the speaker only provided punctate gaze information, looking to the target of her referential utterance briefly after naming it and then looking forward towards the camera for the remainder of the naming event. On the basis of the results from Experiments 1 and 2, we predicted that Extended Hold trials would show less developmental differentiation in social attention and be easier to learn from. In contrast, we predicted that following the Brief Look would require rapid reallocation of social attention, and thus that older children would succeed in following the speaker's gaze while younger children failed. Further, we predicted that this difference in gaze-following would produce down-stream differences in learning.

As predicted, on Extended Hold trials, children in all age groups spent a large proportion time looking at the Target object ($\mu_{1-year} = .58$, $\mu_{2-year} = .62$, $\mu_{3-year} = .65$, $\mu_{4-year} = .64$), and age was weakly correlated with looking behavior ($r(288) = .14, p < .05$). In contrast, children spent a much lower proportion of Brief Look events looking at the Target, and this proportion increased across much more across development ($\mu_{1-year} = .10$, $\mu_{2-year} = .12$, $\mu_{3-year} = .21$, $\mu_{4-year} = .25$; $r(287) = .45$, $p < .001$). To test whether these correlations were different, we fit a mixed-effects model predicting looking at the target on these learning trials from cue type (Brief vs. Extended), age, and their interaction. The interaction term was significant, indicating that age predicted more of the variance in looking on Brief Look trials ($\beta_{BriefLook*age} = .03$, $t = 3.64$, $p < .001$).

As before, children in all age groups showed evidence of processing Familiar words at above chance levels (smallest $\mu_{1-year} = .59$, $t(63) = 5.04$, $p < .001$). Children two-years-old and older showed evidence of learning from the Extended Hold trials (smallest $\mu_{2-year} = .61$, $t(66) = 5.70$, $p < .001$), but only the 3- and 4-year-olds learned from the Brief Look trials (smallest $\mu_{3-year} = .58$, $t(67) = 4.25$, $p < .001$). To confirm these analyses, we fit a linear mixed effects model predicting looking at the correct referent on Test trials from children's age and the trial type. Children improved significantly over development ($\beta_{age} = .05$, $t = 13.83$, $p < .001$). Children's Test trial performance also varied significantly across trial types. Relative to novel words encountered in Extended Hold events, children performed better on Familiar words($\beta_{Familiar} = .11$, $t = 8.85$, $p < .001$) and worse on Novel words encountered in Brief Look events ($\beta_{BriefLook} = -.08$, $t = -6.74$, $p < .001$).

These results confirm our first predictions: Children improved in following social cues across developments, and the size of this improvement was larger when the the cue required rapid re-allocation of attention. Children also learn less from these hard-to-follow Brief Looks. We next provide two convergent analyses that confirm our final prediction: Individual differences in looking behavior on learning trials predicted individual differences in learning.

First, we fit a mixed effects model predicting children's looking on test trials from their age, their Familiar word processing, and their looking on learning trials. This model showed significant effects of all three predictors, confirming that individual differences in social information processing predicted individual differences in learning ($\beta_{age} = .05$, $t = 5.08$, $p < .001$; $\beta_{Familiar} = .17$, $t = 2.41$, $p < .05$, $\beta_{learning} = .15$, $t = 6.23$, $p < .001$). This model was not improved by adding social cue type, and predicted significantly more variance than a model in which we included cue type instead of individual learning scores ($\chi^2 = .75$, $p < .001$). Thus, experimentally manipulating the social cue induced individual differences in children's social attention, and these individual differences predicted individual differences in downstream learning.

Finally, we asked whether this relationship varied between the Extended Hold and the Brief Look. Individual differences in children's ability to follow the Extended Hold were marginally correlated with their learning ($r(255) = .11$, $p = .08$). In contrast individual differences in children's success in following the Brief Look trials predicted individual differences in learning from Brief Looks ($r(256) = .32$, $p < .001$). To test whether these correlations were different, we fit a mixed-effects model predicting test accuracy from cue type (Brief vs. Extended), proportion of time looking at the target during learning trials, and their interaction. We found a significant interaction, indicating that variability in children's social attention were more consequential when demands on social information

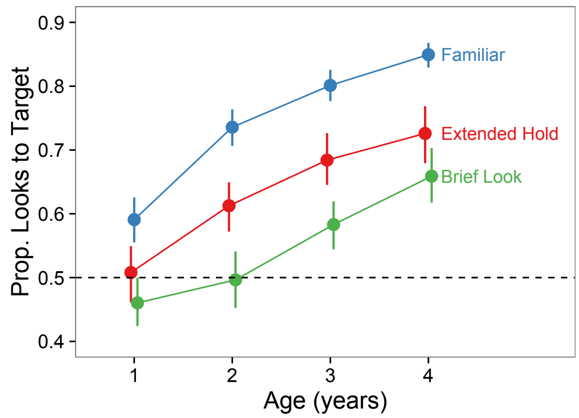

Figure 4. Test trial performance in Experiment 3 for the Familiar words, as well as for Novel words that appeared in Brief Look and in Extended Hold learning events. Error bars show 95% confidence intervals computed by non-parametric bootstrap; points are offset on the horizontal to avoid overplotting.

processing were greater; when the speaker provided only a Brief Look ($\beta_{ExtendedHold*learning} = -.22$, $t = -2.22$, $p < .05$).

Together, these results demonstrate the powerful role of fast social information processing in early word learning. All of the children were able to follow the speaker's social cues when they Extended over the entire naming event, and this supported their learning. In contrast, only the older children were able to follow the Brief Look, and only they were consequently able to learn from it.

Discussion

The early emergence of children's linguistic and communicative capacities is extraordinary. Infants show evidence of following social cues and understanding the communicative function of language by 6 months of age (12, 33). By the same early age, infants can learn about the structure of their languages by tracking the distributional properties of the speech they hear, and even appear to have at least nascent meanings for some words (34, 35). A large body of research in developmental psychology has taken the early emergence of these abilities as evidence for expertise, implicitly or explicitly endorsing the idea that children learn language quickly because they are expert social and distributional information processors.

But early competence is not enough to produce rapid learning. Children must also be able to *perform* these abilities rapidly and robustly in complex real-world settings. We show that this performance has a long developmental trajectory: social information processing improves dramatically over the first five years of life. While one-year-olds were able to follow social gaze occasionally in our experiments, this level of performance did not translate into novel word learning. And even four-year-olds, whose performance was much better, still did not shift their social attention as flexibly as adults.

This social information processing bottleneck also characterized the learning of our sample of children with autism spectrum disorder. Autism is a complex and heterogeneous disorder, affecting different children to different degrees, and in different ways. Nonetheless, one of the core deficits appears to be a disorder of social information processing. This deficit was manifest in our data—children with autism spectrum disorder performed substantially lower than typically developing children in following social cues and learning new words. But critically, variability in social information processing among these children predicted variability in learning just as with the typically developing children. We take this as evidence for the generality of the importance of social information to word learning: Successful learners were successful in the same way.

Although these experiments provide evidence for pronounced and continual improvement in children's social information processing, they leave open the question of what is responsible for these changes. One possibility is that these changes in social attention are caused by domain-general developmental changes in attentional control more broadly (36, 37). Alternatively, these children could be refining their domain-specific representations of the visual and temporal structure of conversations, producing better predictions about where speakers will look and reach next (38, 39). In either case, we propose that these changes may play a powerful role in explaining the rapidly accelerating pace of children's language learning. Indeed, while much has been made of the changes in word learning that occur in the first and second years, the rate at which children learn words continues to increase over the third, fourth, and fifth years (40).

Of course, these experiments are not without their limitations. Although we endeavored to recruit a large and developmentally diverse set of participants, the primary data in Experiments 1 and 2 are correlational in nature. Experiment 3 was designed to causally manipulate social information, providing stronger evidence for the correlation observed in the Experiments 1 and 2. In addition, to minimize variability in stimulus presentation across children, we used a set of fixed video stimuli. It is possible that the use of videos actually underestimates children's social information processing and learning (41). However, more recent work has shown that the "video deficit" in learning is ameliorated when children observe reciprocal interactions like the dialogues in Experiment 1 (42).

Finally, to minimize measurement noise, in our experiments we artificially fixed the learning environment for each child and made the child a static, third-person observer. In contrast, in the natural context of learning, input varies considerably across children and across development and is observed from a first-person perspective (25). This variability has meaningful consequences: Children who receive more high-quality language input learn more language (43). But given our findings on changes in speed of social attention, rapid language learning may require more than the right input; children might need the right input at the right time.

If infants' social information processing performance is so poor, why do they learn words so rapidly? One intriguing possibility is that parents may tune their visual and linguistic input to their children's developing social information processing skills (44–46). That is, caregivers may not use Brief Looks to indicate their referents for young children, but instead use Extended Holds (Experiment 3). Perhaps rapid early word learning is not a result of early expertise per se, but instead emerges from the interaction between children's developing processing skills and their caregivers' coordinated linguistic and social input.

Methods

Participants

Data from typically-developing children were collected at the San Jose Children's Discovery Museum, where parents and their children were invited to participate in an experiment investigating children's early word learning after providing informed consent. In Experiment 1, we collected both demographic and eye-tracking data from 349 children in the target age range, of whom 113 were excluded from the final sample for one or more of the following reasons: unacceptable eye-tracker calibration or data ($N = 49$), atypical developmental trajectories ($N = 30$), and less than 75% parent-reported exposure to English ($N = 41$). Our final sample included 238 children (42 1-year-olds (21 girls); 66 2-year-olds (32 girls); 74 3-year-olds (36 girls); and 56 4-year-olds (30 girls)). In Experiment 3, we collected demographic and eye-tracking data from 425 children in the target age range, of whom 200 were excluded from the final sample for one or more of the following reasons: unacceptable eye-tracker calibration or data ($N = 118$), atypical developmental trajectories ($N = 36$), and less than 75% parent-reported exposure to English ($N = 46$). Our final sample included 225 children, ages 1—5 (61 1-year-olds (25 girls); 57 2-year-olds (30 girls); 51 3-year-olds (25 girls); and 50 4-year-olds (24 girls)).

Data from children with ASD were collected at Stanford University. Children were primarily recruited through the Autism and Developmental Disorders Research Registry, and by flyers posted in the Autism and Developmental Disorders Clinic. Children with a diagnostic history of ASD underwent a comprehensive diagnostic evaluation to determine the accuracy of the previous diagnosis based on DSM-5 criteria, which was confirmed with research diagnostic methods. These diagnostic methods included the ADI-R (47, 48) and the Autism Diagnostic Observation Schedule—Generic (ADOS-G) (49, 50). Exclusion criteria included: 1) a genetic, metabolic, or infectious etiology for ASD on the basis of medical history, neurological history, and available laboratory testing for inborn errors of metabolism and chromosomal analysis; and 2) a DSM-5 diagnosis of any severe mental disorder (e.g. schizophrenia or bipolar disorder). We collected demographic and eye-tracking data from 51 1–7-year old-children, of whom 10 were excluded for unacceptable eye-tracker calibration. The final sample comprised 41 children (M_{age} = 4.22-years, range = (2.24–7.97-years), 4 girls).

Adult participants in Experiment 1 were 17 Stanford undergraduates who participated in exchange for course credit. Informed consent was obtained from parents of all children, and from all adult participants, before experiments began.

Stimulus and Design

After an initial eye-tracker calibration phase, children and adults in both experiments watched a ∼6 minute video. Each video presented two kinds of trials. Learning trials consisted of videos in which speakers seated at a table with two toys provided social cues and labeled one of the toys. Test trials showed pictures of two objects on a black background while a voice asked the participant to look at one of them (as in 20).

In addition, each video included a re-calibration stimulus in which a small, brightly colored object move around the screen. These phases of the videos were used to correct the calibrations estimated at the beginning of the experiment (see 51). Finally, videos contained a small number of filler trials consisting of engaging pictures or videos designed to maintain participants' attention.

Learning trials varied across the Experiments. In Experiment 1 and 2, these learning trials consisted of two dialogues in which two speakers sat at a table together, and one referred to each of the two novel toys whose names were taught in the experiment ("toma" and "fep"). The other two learning trials were monologues in which one of the speakers sat at a table alone, with one of the novel toys and one familiar toy and referred to each in turn. Each naming event consisted of six events: first a naming phrase, then a look at the object accompanied by a comment, a second naming, a reach for the object, and finally a demonstration of the object's function accompanied by a third naming. Each novel object was named nine times in total over the course of the video. Participants watched all of the learning trials before they began the test trials. Eight of these test trials tested familiar objects (e.g. dog/car or lamp/carrot); the other eight paired the two novel objects. Naming phrases were of the form "Look at the [car/fep]! Do you see it?" and were spoken by the actors in the video.

In Experiment 3, we simplified the learning trials. All naming sequences were monologues and both toys on the table were novel. Four of the naming sequences were Extended Hold trials, in which the speaker reached for and interacted with one of the two toys while describing its function and producing its name three times. The other four were Brief Look trials in which the speaker produced the same kind of description but indicated the target toy only with a brief look after first producing its label. Brief Look and Extended Hold trials used a distinct but consistent set of two toys, and the same toy was consistently either the target or the competitor on each trial type. Because children in Experiment 1 sometimes lost interest in these videos before reaching the test trials, in Experiment 3 test trials were interspersed with learning trials so that at least some test data could be acquired from each child. In total, Experiment 3 contained 20 test trials. Eight of these test familiar objects as in Experiments 1 and 2. Eight paired the named objects from the learning trials against their foils from the learning trials, four for the object named in Brief Look trials, and four for the object named in Extended Hold trials. The remaining four test trials paired the two named objects against each other, with children being asked to find each two times.

Data Analysis

In all experiments, raw gaze data were transformed before statistical analysis. First to ensure appropriate precision in area-of-interest analyses, infants' calibrations were corrected and verified via robust regression (described in 51), and calibration corrections were assessed by two independent coders ($\kappa_{Exp1} = .8$, $\kappa_{Exp2} = .87$, $\kappa_{Exp3} = .77$). Children whose calibrations could not be verified and corrected were excluded from further analyses.

Analyses were performed use an Area of Interest (AOI) approach. On learning trials, AOIs were hand-coded frame-by-frame for the speakers' faces and for the two on-screen objects. On test trials, these AOIs corresponded to the screen positions of the two alternatives. To use standard statistical analyses, we transformed the timecourse data to looking proportions within relevant windows. On test trials, the start of this window was set to 500ms after the point of disambiguation—the onset of the target label. The end of the window was set to the length of the shortest test trial for accurate averaging. This was 4s for Experiments 1 and 2, and 4.5s for Experiment 3. These windows were chosen to give children sufficient time to process the label, and to maximize signal (see Fig. 2).

In Experiments 1 and 2, Learning trials contained six distinct phases (Fig. 1): a baseline period ($2s$), name 1 to look (M = 1.7s), look to name 2 (M = 2s), name 2 to initiation of reach (M= 4.8s), reach to point of contact ($M = .8s$), and after contact with the toy (M = 1s). Proportion of looks to the target AOI were computed in each of these windows. In Experiment 3, learning trials were designed to separate demands on social attention across rather than within trials. In Experiment 3 we computed proportion of looking of the entirety of learning trials.

Due to occasional bouts of inattention, eye-gaze data were not available for all children during all portions of the experiment. To correct for statistical errors introduced by averaging over small windows, data from individual learning and test trials were excluded from analysis if less than 50% of the window contained eye-tracking data (regardless of where children were looking). Second, if more than 50% of the trials for a given participant were excluded in this manner, all of the remaining trials were dropped as well. All data and code are freely available through the Open Science Framework (`https://osf.io/kjr98/`), and a public GitHub repository (`http://github.com/dyurovsky/refword`).

Acknowledgments

We gratefully acknowledge the parents and children who participated in this research and the staff of the San Jose Children's Discovery Museum for their collaboration on this line of research. This work was supported by NIH NRSA F32HD075577 to DY, grants from the John Merck Scholars program, the Stanford Child Health Research Initiative to MCF, and a grant from the Mosbacher Family Fund for Autism Research at Stanford University.

References

Goulden, R., Nation, P. & Read, J. How large can a receptive vocabulary be? *Applied Linguistics* **11**, 341–363 (1990).

Lidz, J., Waxman, S. & Freedman, J. What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition* **89**, 295–303 (2003).

Bulf, H., Johnson, S. P. & Valenza, E. Visual statistical learning in the newborn infant. *Cognition* **121**, 127–132 (2011).

Shukla, M., White, K. S. & Aslin, R. N. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences* **108**, 6038–6043 (2011).

Bates, E. *et al.* Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* **21**, 85–123 (1994).

Carey, S. & Bartlett, E. Acquiring a single new word. *Papers and Reports on Child Language Development* **15**, 17–29 (1978).

Yu, C. & Smith, L. B. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* **18**, 414–420 (2007).

Frank, M. C., Goodman, N. & Tenenbaum, J. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* **20**, 578–585 (2009).

Scaife, M. & Bruner, J. S. The capacity for joint visual attention in the infant. *Nature* (1975).

Baldwin, D. A. Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology* **29**, 832–843 (1993).

Hollich, G. J., Hirsh-Pasek, K. & Golinkoff, R. M. Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monographs of the Society for Research in Child Development* (2000).

Senju, A., Csibra, G. & Johnson, M. H. Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition* **108**, 303–319 (2008).

Corkum, V. & Moore, C. The origins of joint visual attention in infants. *Developmental Psychology* **34**, 28 (1998).

Brooks, R. & Meltzoff, A. N. The development of gaze following and its relation to language. *Developmental Science* **8**, 535–543 (2005).

Csibra, G. & Gergely, G. Natural pedagogy. *Trends in Cognitive Sciences* **13**, 148–153 (2009).

Sokol, S. Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Research* **18**, 33–39 (1978).

Banks, M. S. The development of visual accommodation during early infancy. *Child Development* 646–666 (1980).

Forssberg, H., Eliasson, A., Kinoshita, H., Johansson, R. & Westling, G. Development of human precision grip i: Basic coordination of force. *Experimental Brain Research* **85**, 451–457 (1991).

Thelen, E. Motor development: A new synthesis. *American Psychologist* **50**, 79 (1995).

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A. & McRoberts, G. W. Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science* **9**, 228–231 (1998).

Adolph, K. E. & Robinson, S. R. Motor Development. In *Handbook of child psychology and developmental science Vol. Cognitive processes*, 114–157 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015).

Adolph, K. E. *et al.* How Do You Learn to Walk? Thousands of Steps and Dozens of Falls per Day. *Psychological Science* **23**, 1387–1394 (2012).

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N. & Beall, A. C. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception* **37**, 1443–1457 (2008).

Vida, M. D. & Maurer, D. The development of fine-grained sensitivity to eye contact after 6years of age. *Journal of Experimental Child Psychology* **112**, 243–256 (2012).

Yu, C. & Smith, L. B. Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS ONE* **8**, e79659 (2013).

Dempster, F. N. Memory span: Sources of individual and developmental differences. *Psychological Bulletin* **89**, 63 (1981).

Kail, R. Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin* **109**, 490 (1991).

Gathercole, S. E., Pickering, S. J., Ambridge, B. & Wearing, H. The structure of working memory from 4 to 15 years of age. *Developmental Psychology* **40**, 177–190 (2004).

Fernald, A., Zangl, R., Portillo, A. L. & Marchman, V. A. Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing* 113–132 (2008).

Smith, L. B. & Yu, C. Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* **106**, 1558–1568 (2008).

Baron-Cohen, S., Baldwin, D. & Crowson, M. Do children with autism use eye-direction to infer linguistic reference. *Child Development* **68**, 48–57 (1997).

Leekam, S. R., Hunnisett, E. & Moore, C. Targets and cues: Gaze-following in children with autism. *Journal of Child Psychology and Psychiatry* **39**, 951–962 (1998).

Vouloumanos, A., Martin, A. & Onishi, K. H. Do 6-month-olds understand that speech can communicate? *Developmental Science* **17**, 872–879 (2014).

Thiessen, E. D. & Saffran, J. R. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* **39**, 706–716 (2003).

Bergelson, E. & Swingley, D. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* **109**, 3253–3258 (2012).

Rueda, M. R. & Rothbart, M. K. Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences* **102**, 14931–14936 (2005).

Smith, L. B. It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist* **68**, 618–629 (2013).

Acheson, D. J. & MacDonald, M. C. Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin* **135**, 50–68 (2009).

Krogh-Jespersen, S., Liberman, Z. & Woodward, A. L. Think fast! The relationship between goal prediction speed and social competence in infants. *Developmental Science* **18**, 815–823 (2015).

Bloom, P. *How children learn the meanings of words* (MIT press Cambridge, MA, 2000).

Anderson, D. R. Television and Very Young Children. *American Behavioral Scientist* **48**, 505–522 (2005).

O'Doherty, K. *et al.* Third-party social interaction and word learning from video. *Child Development* **82**, 902–915 (2011).

Weisleder, A. & Fernald, A. Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science* **24**, 2143–2152 (2013).

Snow, C. E. Mothers' speech to children learning language. *Child Development* **43**, 549–565 (1972).

Gogate, L. J., Bahrick, L. E. & Watson, J. D. A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development* **71**, 878–894 (2000).

Brand, R. J., Baldwin, D. A. & Ashburn, L. A. Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science* **5**, 72–83 (2002).

Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Devlopmental Disorders* **24**, 659–685 (1994).

Le Couter, A. *et al.* Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Devlopmental Disorders* **19**, 363–387 (1989).

Lord, C., Rutter, M., DiLavore, P. C. & Risis, S. *Autism diagnostic observation schedule: Manual* (Los Angeles, CA: Western Psychological Services., 1999).

Lord, C., Risi, S., Lambrecht, L. & Cook Jr, E. H. The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Devlopmental Disorders* 205–223 (2000).

Frank, M. C., Vul, E. & Saxe, R. Measuring the development of social attention using free-viewing. *Infancy* **17**, 355–375 (2012).