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

Daniel Yurovsky [*], Anna Wade [†], Allison M Kraus [*] , Grace W. Gengoux [‡], Antonio Hardan [‡] , and Michael C. Frank [*]

[*] Department of Psychology, Stanford University, [†] School of Medicine, University of California, San Francisco, and [‡] School of Medicine, Stanford University

How do children learn new words so rapidly? To begin learning the meaning of a word, a child must determine if this word is being used to refer to something in the immediate physical context. Social cues—like the eye-gaze of a helpful speaker—are powerful information sources for resolving this problem. Studies of children's gaze following have generally focused on its early emergence, demonstrating that even 3-month-olds show measurable success in following social gaze. This early competence in using social information has often been taken to explain their rapid learning. We show, however, that this early competence has a long that this 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, predicting how well children learn new words over the same time period. Finally, we show that the same bottleneck exists in children diagnosed with autism spectrum disorder. These results describe a route by which increases in social expertise can lead to changes in language learning ability and highlight the dependence of developmental outcomes on not just the existence of particular competencies, but on their proficient use in complex contexts.

language acquisition | word learning | social cues | cognitive development

Children's first five years are a time of rapid developmental 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 her first birthday to producing complex, grammatical sentences only a few years later. The remarkable rate and breadth of this developmental change have galvanized psychologists and neuroscientists to search for the mechanisms that allow children to learn so much so quickly. Perhaps because of the fundamentally discrete nature of the units of language itself, much of the work in this search has focused specifically on the origins of these important mechanisms (but c.f. 1, 2). Reviews of language development are often a list of milestones: e.g., infants acquire the ability to use sequential statistics to segment speech at 8 months (3), begin to use co-occurrence statistics to learn word-object mappings at 12 months (4), and begin to use syntax to infer the meanings around 24 months (5). New exciting work then is often a demonstration of some competence earlier than previously observed (6). However, learning outside the laboratory is controlled not by early competence, but by proficient use these mechanisms in complex natural environments.

Here we take as a case study children's use of social information to infer the meanings of new words. Over the first five years, children acquire a productive vocabulary of approximately 4000–5000 words (7). Though these words belong to many grammatical categories, a large proportion are concrete nouns (8). 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 (9–11)? A powerful source of information for resolving this uncertainty is available in the social cues provided by the speaker: e.g., where is the speaker looking? Consequently, a large body of research has accumulated documenting young infants' abilities to track a speaker's social cues and use them to infer the target of her reference (e.g. 12–15).

This research has focused almost exclusively 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. 16–18). This research strategy stands in contrast to work in domains like visual and motor development, in which researchers have sought to measure continuous improvement in children's abilities as they develop (19–22). As these researchers, and numerous engineers and roboticists have repeatedly discovered, there is no mistaking the ability to take a single step and the ability to run a marathon (23). We argue, however, that the study of language development has made exactly this mistake.

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 (24–26). To use social cues like eye gaze to resolve the uncertainty about a speakers' intended referent, children must rapidly process the auditory and visual information they are receiving. Competence 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 the information in memory (27–29).

In three experiments, we show that the developing ability to rapidly and flexibly allocate attention in social interactions is a bottleneck

---

Significance

The study of language development has historically focused on pinpointing children's earliest points of competence across domains. However, learning outside the laboratory is controlled by proficient use of the skills in these domains. While infants can follow social cues like eye-gaze in their first year, the speed and fidelity of this skill improves dramatically over the first five years. The problem of tracking social information is far from resolved for young infants; it remains a bottleneck on word learning for typically developing children into the preschool years and and does so for children with autism spectrum disorder as well. This work highlights the importance of studying not just origins, but also the developmental trajectories of higher-level cognitive processes.

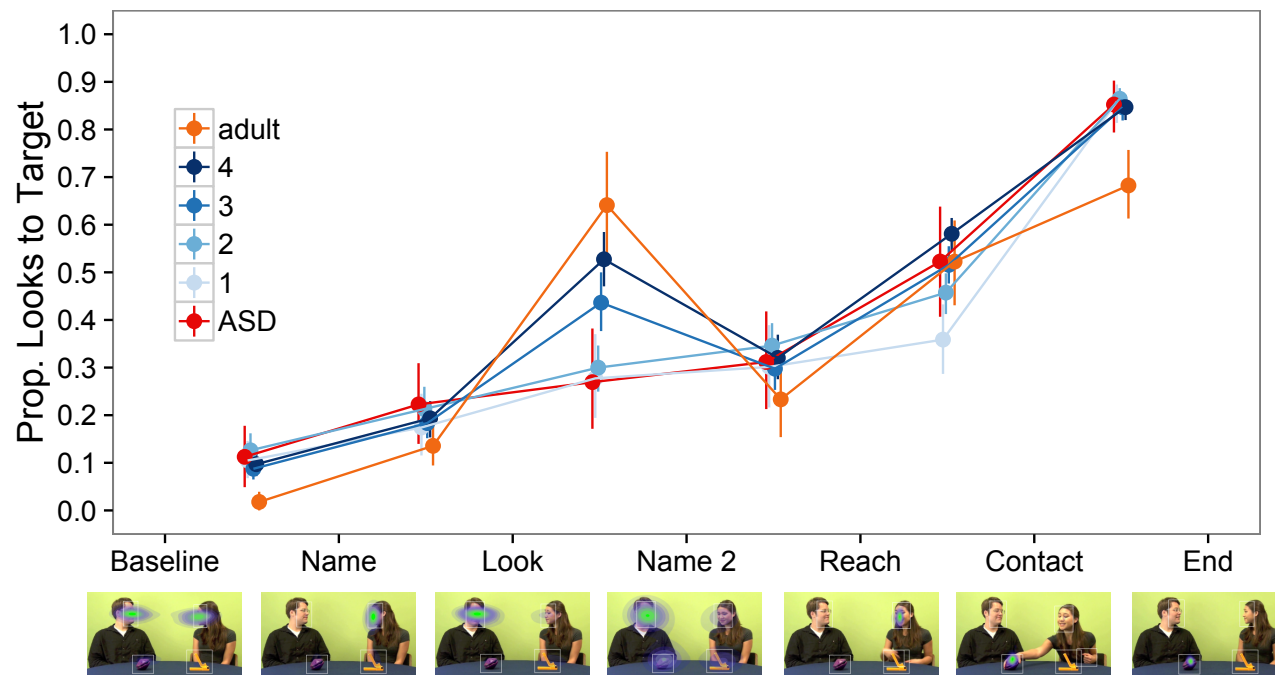**Reserved for Publication Footnotes**

Figure 1. (bottom) 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)

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 intended 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 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 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 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.

Experiment 1: Social Attention and Word Learning

Experiment 1 was designed to estimate the developmental trajectory of children's online social information process in complex interactions, and to ask whether this processing had downstream conse-

quences 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 = 7.66, p < .001 ; $\beta_{reach-contact}$ = .08, t = 5.83, p < .001). While the youngest children were only occasionally able to quickly follow the speaker's social cues, older children and adults were much more consistent. Thus, the ability to quickly and reliably process social information online 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 (1, 30). On Test trials, children saw two toys 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").
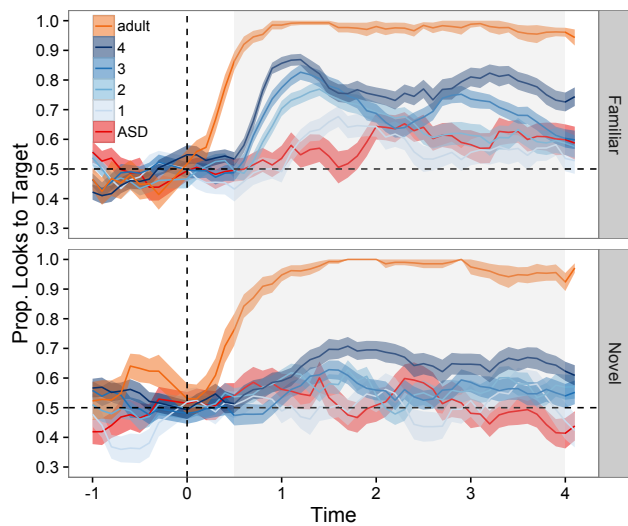
Figure 2. Children's and adults' looks to the Target and Competitor objects over the course of Test trials for typically developing children and adults (Experiment 1) and children with ASD (Experiment 2). Lines show age-group means, and shaded regions show standard errors of these means. The light gray rectangle shows the window over which looking proportions were computed for subsequent statistical analyses.

On other trials, the target was a familiar whose label is typically already in known to 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 $\mu_{2-year} = .56, t(53) = 3.89, p < .001$). A linear model showed that children performed better on Test trials over development ($\beta_{age} = .07, z = 7.93, p < .001$), and that the effect of age was greater for Familiar than Novel trials ($\beta_{age*novel} = -.012$, $z = -2.42, p < .05$, 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 ($r_{age}(197) = .29, p < .001$); $r_{familiar}(197) = .20, p < .01$; $r_{look-name2}(192) = .34, p < .001$; $r_{reach-contact}(194) = .16, p < .05$). Which of these factors 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 ($\beta_{look-name2} = .14, z = 3.31, p < .01$), although age was marginal ($\beta_{age} = .02, z = 1.66, p = .1$). 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 ($\beta_{look-name2} = .12, z = 2.94, p < .01$).

Thus, age-related improvements in social attention, whatever their cause, are a critical bottleneck on learning from naming events throughout early childhood. Children who were unable to quickly 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.

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
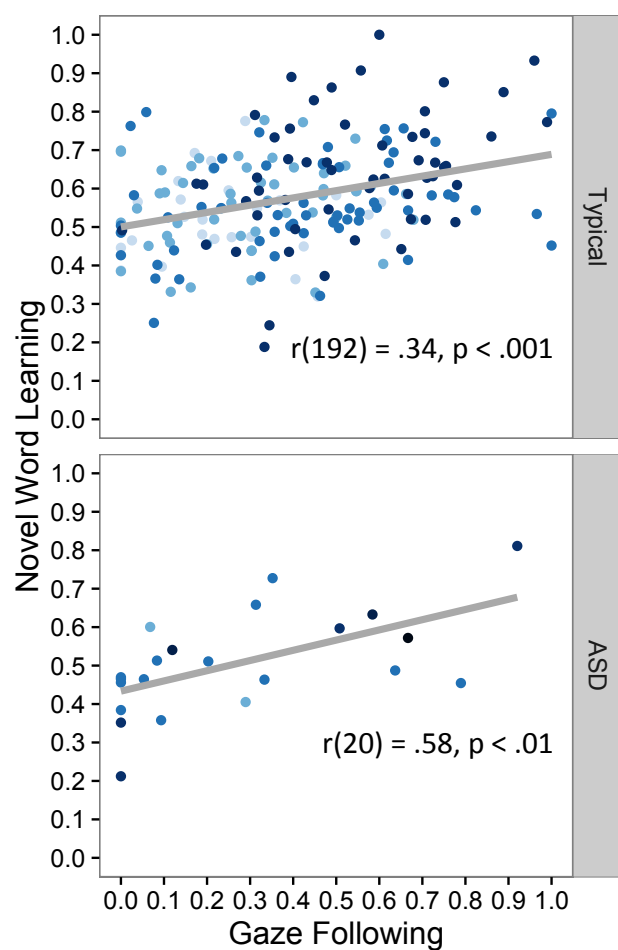

Figure 3. Correlations between gaze-following on Learning trials and accuracy on Novel Test trials for typically developing children in Experiment 1 and children with autism spectrum disorder (ASD) in Experiment 2. Darker colored points indicate older children.

the informative social cues. Second, children could learn these meanings across the interactions, by using co-occurrence statistics between the words and objects. While typically developing children's social information processing predicted a large chunk of the variance in their word learning, it is possible that atypically developing children would use a different strategy.

We tested a group of 40 2–8-year-old children diagnosed with autism spectrum disorder (ASD) who, due to deficits in social information processing, might be expected to learn predominantly from the cross-situational statistics instead. 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($r(27) = .12, p = .54$), following the speaker's social gaze was highly correlated with Novel word learning ($r(20) = .58, p < .01$), just as in the typically developing sample (Fig. 4). We again fit a linear 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 ($\beta_{familiar} = .50, z = 2.40$, $p < .05$) and gaze following ($\beta_{look-name2} = .25, z = 2.96, p < .01$).

Thus, social information processing is a bottleneck on word learning for children with autism spectrum disorder, just as it is for typi-

cally 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 learn the Novel words did so by following social cues.

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.

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 downstream differences in learning.

As predicted, on Extended Hold trials, children spent a large proportion of the naming events looking at the Target across age groups ($\mu_{1-year} = .57$, $\mu_{2-year} = .60$, $\mu_{3-year} = .63$, $\mu_{4-year} = .60$), and age was not significantly correlated with looking behavior ($r(290) = .08$, $p = .15$). In contrast, children spent a much lower proportion of Brief Look events looking at the Target, and this proportion increased across age ($\mu_{1-year} = .09$, $\mu_{2-year} = .09$, $\mu_{3-year} = .16$, $\mu_{4-year} = .17$; $r(288) = .27$, $p < .001$).

As before, children in all age groups showed evidence of processing Familiar words at above chance levels (smallest $\mu_{1-year} = .59$, $t(63) = 5.09$, $p < .001$). Children two-years-old and older showed evidence of learning from the Extended Hold trials (smallest $\mu_{2-year} = .63$, $t(66) = 5.43$, $p < .001$), but only the 3- and 4-year-olds learned from the Brief Look trials (smallest $\mu_{3-year} = .58$, $t(68) = 3.66$, $p < .001$). To confirm these analyses, we fit a 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 ($\beta_{age} = .08$, $t = 19.34$, $p < .001$). 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 ($\beta_{Familiar} = .09$, $t = 6.64$, $p < .001$) and worse on Novel words encountered in Brief Look events ($\beta_{BriefLook} = -.09$, $t = -6.48$, $p < .001$).

To succeed on Test trials, children needed to succeed at two tasks: 1. Process social information available in Learning trials to disambiguate naming events and discover the speaker's intended referent, and 2. Process the language they heard at Test, and use it to retrieve the correct mapping between the Novel word and the Target referent. The Extended Hold events were designed to reduce demand on children's online social information processing (1), rendering language processing and memory retrieval (2)the only bottleneck on Test trials. We leveraged this relationship to perform an additional test of our primary hypothesis: We fit a mixed effects model predicting children's Novel Test performance from their age, their Familiar word processing, and the naming event type (Extended Hold vs. Brief Look), as well as the interaction of Familiar word processing and Test type. This model showed a significant effect of age ($\beta_{age} = .05$, $t = 4.83$,
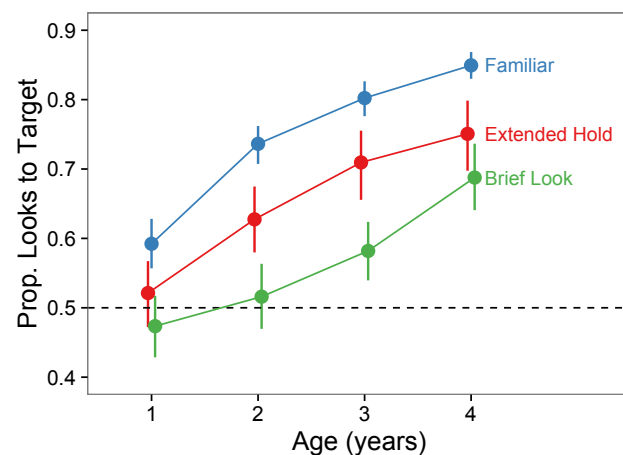

Figure 4. Test trial performance in Experiment 3 for the Familiar words, as well as for Novel words that appeared in Brief Look and in Extended Hold learning events. Error bars show 95% confidence intervals computed by non-parametric bootstrap; points are offset on the horizontal to avoid overplotting.

$p < .001$), as well as a significant interaction between Familiar word processing and naming event type ($\beta_{ExtendedHold*Familiar} = .21$, $t = 3.47$, $p < .001$). That is, Familiar word processing was a better predictor of children's success on Extended Hold trials, where it was the primary bottleneck.

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 (15, 31). 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 (6, 32). Children also acquire language at an impressive rate. A large body of research in developmental psychology has taken the first to be the *cause* of the second: Children learn language so 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. The experiments in this paper 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 at above-chance levels, this level of performance did not translate into novel word learning. Even four-year-olds, whose performance was much better, still did not shift their social attention or consequently learn words as well 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 was borne out in our data—children with autism spectrum disorder significantly underperformed age-matched 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 generality of social information processing to word learning: Successful learners were successful in the same way.

Although these experiments provide strong 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 (33, 34). 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 (35, 36). In either case, we propose that these changes have 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 only increases over the third, fourth, and fifth years (37).

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 (38). However, more recent work has shown that the "video deficit" in learning is ameliorated when children observe the kinds of reciprocal interaction that characterize the dialogues in Experiment 1. In addition, while we endeavored to construct naturalistic stimuli that represent the kinds of interactions characteristic of children's language learning environments, they are far from exhaustive. In order to minimize measurement noise, we artificially fixed the learning environment for each child, but the natural learning environment varies considerably across children and across development. And this variability may be precisely the cause of rapid early word learning.

If infants' social information processing performance is so poor, why do they learn words so rapidly? One intriguing possibility is that parents may fine-tune their visual and linguistic input to their children's developing social information processing skills (39–41). We thus propose that rapid early word learning may result not from children's early expertise in social information processing, for they are far from expert. Instead, rapid early word learning emerges from the interaction between children's developing processing skills and their caregivers' coordinated linguistic and social input.

Materials and 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 ($N = 52$), atypical developmental trajectories ($N = 30$), and less than 75% parent-reported exposure to English ($N = 40$). Our final sample included 236 children (41 1-year-olds (21 girls); 66 2-year-olds (32 girls); 73 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 201 were excluded from the final sample for one or more of the following reasons: unacceptable eye-tracker calibration or data ($N = 96$), atypical developmental trajectories ($N = 36$), and less than 75% parent-reported exposure to English ($N = 48$). Our final sample included

224 children, ages 1 – 5 (60 1-year-olds (24 girls); 57 2-year-olds (29 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 with ASD and their siblings 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 (42, 43) and the Autism Diagnostic Observation Schedule – Generic (ADOS-G) (44, 45). 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 such as schizophrenia and bipolar disorder. We demographic and eye-tracking data from 51 1–7-year old-children, of whom 11 were excluded for unacceptable eye-tracker calibration. The final sample comprised 40 children (M_{age} = 4.26-years, range = (2.24–7.97-years), 4 girls).

Adult participants in Experiment 1 were 17 Stanford undergraduate students 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 1).

In addition, each video included a re-calibration stimulus which participants saw a small, brightly colored stimulus move around the screen. These phases of the videos were used to correct the calibrations estimated at the beginning of the experiment (see 46). Finally, videos contained a small number of filler trials in which participants saw engaging pictures or videos designed to keep maintain their attention for the duration of the video.

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 went through several transformations before statistical tests were computed. First to ensure appropriate precision in region-of-interest analyses, infants' calibrations were corrected and verified via robust regression (described in 46), and calibration corrections were assessed by two independent coders ($\kappa = FIXME$). Children whose calibrations could not be verified and corrected were excluded from further analyses.

Second, all analyses were performed use an Area of Interest (AOI) approach. On learning trials, AOIs were hand-coded frame-by-frame for each of the speakers' faces and for the two on-screen objects. On test trials, these AOIs corresponded to the static on-screen positions of the two alternatives. To use standard statistical analyses, we transformed the timecourse data to proportions of looking 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 all 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 of exactly 2 s, name 1 to look (M = 1.7 s), look to name 2 (M = 2.0 s), name 2 to initiation of reach (M= 4.8 s), initiation of reach to point of contact (M = .8 s), and after contact with the object (M = 1 s). 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 23 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 analysis code are freely available at a public GitHub repository at `http://github.com/dyurovsky/refword`.

1. Fernald A, Pinto JP, Swingley D, Weinbergy A, McRoberts GW (1998) Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychol Sci* 9:228–231.
2. Port RF, Leary AP (2005) Against formal phonology. *Language* pp. 927–964.
3. Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
4. Smith LB, Yu C (2008) Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106:1558–1568.
5. Gertner Y, Fisher C, Eisengart J (2006) Learning words and rules abstract knowledge of word order in early sentence comprehension. *Psychol Sci* 17:684–691.
6. Bergelson E, Swingley D (2012) At 6–9 months, human infants know the meanings of many common nouns. *P Natl Acad Sci USA* 109:3253–3258.
7. Goulden R, Nation P, Read J (1990) How large can a receptive vocabulary be? *App Linguist* 11:341–363.
8. Bates E et al. (1994) Developmental and stylistic variation in the composition of early vocabulary. *J Child Lang* 21:85–123.
9. Carey S, Bartlett E (1978) Acquiring a single new word. *Pap R Child* 15:17–29.
10. Yu C, Smith LB (2007) Rapid word learning under uncertainty via cross-situational statistics. *Psychol Sci* 18:414–420.
11. Frank MC, Goodman N, Tenenbaum J (2009) Using speakers' referential intentions to model early cross-situational word learning. *Psychol Sci* 20:578–585.
12. Scaife M, Bruner JS (1975) The capacity for joint visual attention in the infant. *Nature*.
13. Baldwin DA (1993) Early referential understanding: Infants' ability to recognize referential acts for what they are. *Dev Psychol* 29:832–843.
14. Hollich GJ, Hirsh-Pasek K, Golinkoff RM (2000) Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monogr Soc Res Child*.
15. Senju A, Csibra G, Johnson MH (2008) Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition* 108:303–319.
16. Corkum V, Moore C (1998) The origins of joint visual attention in infants. *Dev Psychol* 34:28.
17. Brooks R, Meltzoff AN (2005) The development of gaze following and its relation to language. *Developmental Sci* 8:535–543.
18. Csibra G, Gergely G (2009) Natural pedagogy. *Trends Cogn Sci* 13:148–153.
19. Sokol S (1978) Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Res* 18:33–39.
20. Banks MS (1980) The development of visual accommodation during early infancy. *Child Dev* pp. 646–666.
21. Forssberg H, Eliasson A, Kinoshita H, Johansson R, Westling G (1991) Development of human precision grip i: basic coordination of force. *Exp Brain Res* 85:451–457.
22. Thelen E (1995) Motor development: A new synthesis. *Am Psychol* 50:79.
23. Adolph KE, Robinson SR (2015) Motor Development in *Handbook of child psychology and developmental science Vol. Cognitive processes*. (John Wiley & Sons, Inc., Hoboken, NJ, USA), pp. 114–157.
24. Loomis JM, Kelly JW, Pusch M, Bailenson JN, Beall AC (2008) Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception* 37:1443–1457.
25. Vida MD, Maurer D (2012) The development of fine-grained sensitivity to eye contact after 6years of age. *J Exp Child Psychol* 112:243–256.
26. Yu C, Smith LB (2013) Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS ONE* 8:e79659.
27. Dempster FN (1981) Memory span: Sources of individual and developmental differences. *Psychol Bull* 89:63.
28. Kail R (1991) Developmental change in speed of processing during childhood and adolescence. *Psychol Bull* 109:490.
29. Gathercole SE, Pickering SJ, Ambridge B, Wearing H (2004) The structure of working memory from 4 to 15 years of age. *Developmental psychology* 40:177–190.
30. Fernald A, Zangl R, Portillo AL, Marchman VA (2008) Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing* pp. 113–132.
31. Vouloumanos A, Martin A, Onishi KH (2014) Do 6-month-olds understand that speech can communicate? *Developmental Sci* 17:872–879.
32. Thiessen ED, Saffran JR (2003) When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev Psychol* 39:706–716.
33. Rueda MR, Rothbart MK (2005) Training, maturation, and genetic influences on the development of executive attention. *P Natl Acad Sci USA* 102:14931–14936.
34. Smith LB (2013) It's all connected: Pathways in visual object recognition and early noun learning. *Am Psychol* 68:618–629.
35. Acheson DJ, MacDonald MC (2009) Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychol Bull* 135:50–68.
36. Krogh-Jespersen S, Liberman Z, Woodward AL (2015) Think fast! The relationship between goal prediction speed and social competence in infants. *Developmental Sci* 18:815–823.
37. Bloom P (2000) *How children learn the meanings of words*. (MIT press Cambridge, MA).
38. Anderson DR (2005) Television and Very Young Children. *Am Behav Sci* 48:505–522.
39. Snow CE (1972) Mothers' speech to children learning language. *Child Dev* 43:549–565.
40. Gogate LJ, Bahrick LE, Watson JD (2000) A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Dev* 71:878–894.
41. Brand RJ, Baldwin DA, Ashburn LA (2002) Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Sci* 5:72–83.
42. Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24:659–685.
43. Le Couter A et al. (1989) Autism diagnostic interview: A standardized investigator-based instrument. *J Autism Dev Disord* 19:363–387.
44. Lord C, Rutter M, DiLavore PC, Risis S (1999) *Autism diagnostic observation schedule: Manual*. (Los Angeles, CA: Western Psychological Services.).
45. Lord C, Risi S, Lambrecht L, Cook Jr EH (2000) The Autism Diagnostic Observation Schedule|Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* pp. 205–223.
46. Frank MC, Vul E, Saxe R (2012) Measuring the development of social attention using free-viewing. *Infancy* 17:355–375.