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

Submitted to Proceedings of the National Academy of Sciences of the United States of America

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 been concerned with its developmental origins, demonstrating measurable success in early infancy. We show that this ability has a long developmental trajectory, however: Slow, continuous improvements in speed of social information processing occur over the course of the first five years of life. This developing ability is a significant bottleneck on early word learning, predicting changes in children's learning of 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. Perhaps because of the fundamentally discrete nature of the units of language itself, much work in language acquisition has focused specifically on the origins of these important abilities (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). This research strategy stands in contrast to work in domains like visual and motor development, in which researchers have sought to measure continuous changes in children's abilities as they develop (7, 8, 9, 10).

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 (11). Though these words belong to many grammatical categories, a large proportion are concrete nouns (12). 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 (13, 14, 15)? 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. 16, 17, 18, 19).

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. 20, 21, 22). Contradicting this assumption, however, some work has argued that rapid gaze-following may occur relatively rarely in natural interactions and may be difficult even for older children and adults under some circumstances (23, 24, 25).

There are more basic reasons to assume that children's use of social information in word learning would have a protracted developmental trajectory, as well. 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. To learn from a given situation, learners must rapidly process the auditory and visual information they are receiving. Consequently, competence is not enough: the referential uncertainty problem should remain a problem in proportion to a child's developing ability to control her attention, process auditory and visual information, and hold the information in memory (26, 27, 28).

In three experiments, we test the hypothesis that social attention allocation is a bottleneck in early 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 co-occurrence information 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 the hypothesis that those children who were successful in attending to the conversational referent would be the same children who learned and retained the words.

Significance

The study of language development has historically focused on pinpointing children's earliest points of competence in each domain. However, learning outside the laboratory is ultimately controlled by proficient use of these mechanisms. While infants can follow social cues like eye-gaze early in their first year, we show that the speed and fidelity of this process 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 autistic children 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

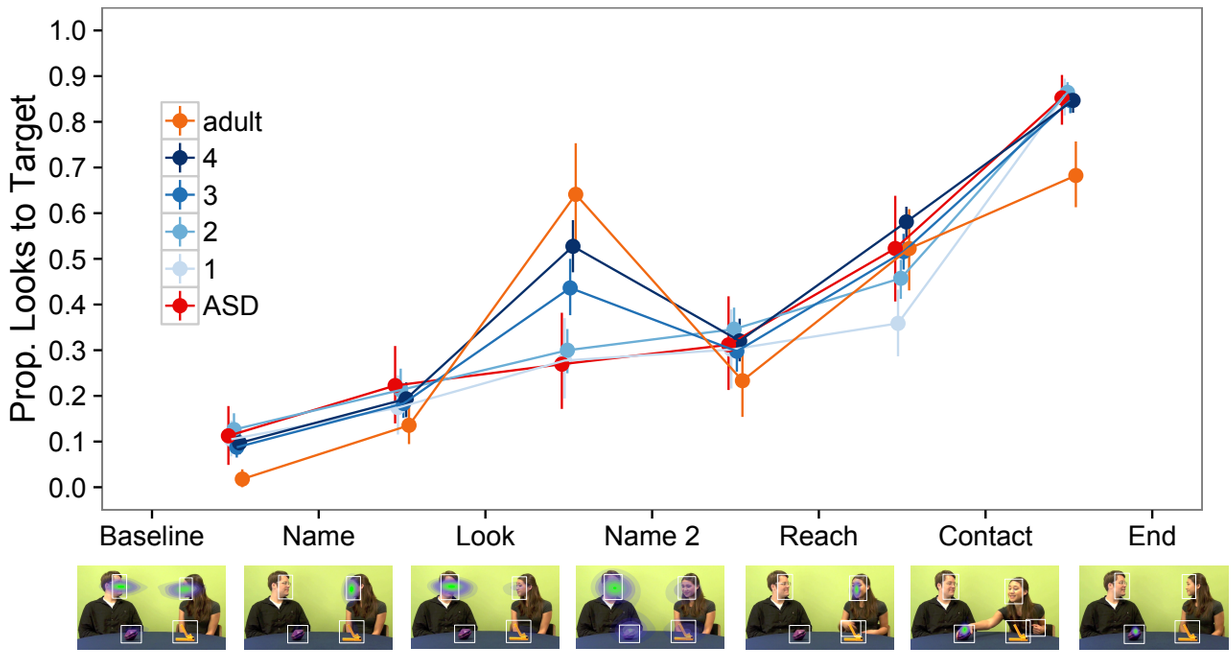

Figure 1. (bottom) Looking to target vs. competitor objects during the learning portion of Experiments 1 (typically-developing children) and 2 (ASD children), plotted by phase of the naming event. Points show means and error bars show 95% confidence interval across participants; points are offset on the horizontal to avoid overplotting. (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 entire participant group's average point of gaze (hotter colors indicate more fixation; scale is constant across frames)

In Experiment 1, we measure developmental changes in social information processing and use 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 to determine if 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 manipulate 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 recruit 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 consequences for word learning. The experiment consisted of two parts: learning and test. During the learning part, 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 participants spent looking at the target referent during the naming sequences in which the novel objects were introduced (Fig. 1, top). We analyzed these 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, nearly all 100% of children were looking to the correct object 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—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 sensitive. Thus, the ability to quickly and reliably process social information online in a naturalistic conversation appears to develop significantly over the course of the first five years, and even further into adulthood.

After watching these learning events, children's retention of the novel words was tested. During the test trials we showed children pairs of toys and labeled one, using a standard looking-while-listening procedure (1, 29). To measure online language comprehension more generally, we also included familiar word trials (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 $\mu_{2-year} = .56$, $t(53) = 3.89$, $p < .001$, Fig. 3). A linear model showed that children performed better on test over development ($\beta_{age} = .07$, $z = 7.93$, $p < .001$), and that the effect of age was greater for familiar than novel trials ($\beta_{age*novel} = -.012$, $z = -2.42$, $p < .05$)

Thus, older children learned more from the same naming events. However, these children also improved on both other measures — quickly following speaker's social gaze and reach, and processing 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

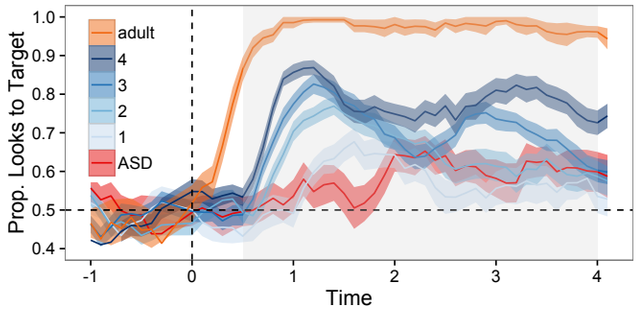

Figure 2. Children's and adults' looking over the course of Familiar test trials in Experiment 1. Lines show group-means and shaded regions show standard errors of these means. The light gray rectangle shows the window over which performance was computed for subsequent statistical analyses.

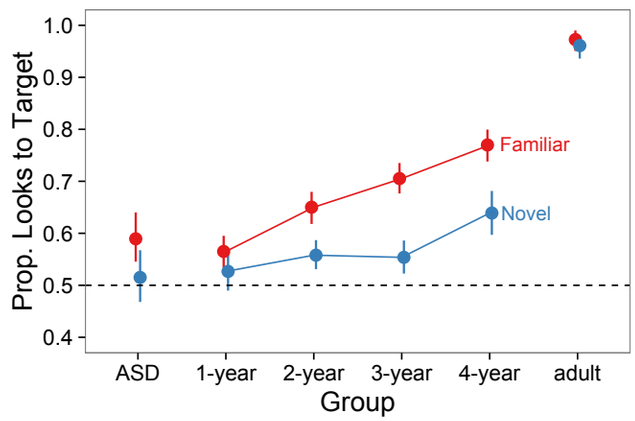

Figure 3. Test trial performance in for children and adults Experiment 1, as well children with ASD in Experiment 2. Colors show Familiar and Novel word trials, and error bars show 95% confidence intervals computed by non-parametric bootstrap. Points are offset on the horizontal to avoid overplotting.

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. Put another way, the age-related changes in word learning that we saw appeared to be accounted for largely by changes in social attention.

Experiment 2: Social Attention in Autistic Children

The naming events in Experiment 1 contained information sufficient to infer the meanings of the novel words at two timescales: (1) within the interactions, by following the informative social cues, and (2) across the interactions, by using co-occurrence statistics between the words and objects. While typically developing children's social information processing predicting a large chunk of the variance in their

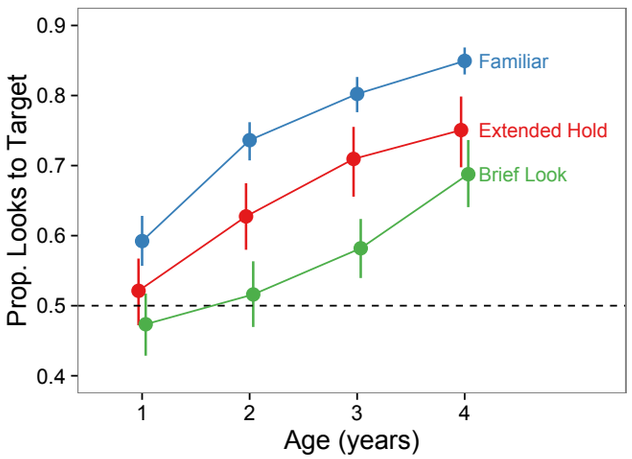

Figure 4. Test trial performance in Experiment 3 for the Familiar words, as well as for Novel words that appeared in Brief Look and in Extended Hold learning events. Error bars show 95% confidence intervals computed by non-parametric bootstrap; points are offset on the horizontal to avoid overplotting.

word learning, it is possible that atypically developing children would use a different strategy.

We tested a group of 40 children diagnosed with Autism Spectrum Disorder 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 learned the novel words, and attended to the referents of familiar words at approximately the same levels (Fig. 3). However, while age was uncorrelated with word learning for the Autistic children ($r(27) = .12$, $p = .54$), following the speaker's gaze was highly correlated with novel word learning ($r(20) = .58$, $p < .01$), just as in the typically developing sample. 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 remains a critical bottleneck on word learning in autistic children, just as in typically developing children.

Experiment 3: Varying Demands on Social Attention

The naming events in Experiment 1 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 analysis showed that fast processing of social gaze was the primary predictor of ultimate word learning. In Experiment 2, we tested this prediction directly.

In Experiment 2, 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 while younger children failed. Further, we predicted that this would produce downstream differences in learning.

As predicted, on Extended Hold trials, children's 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 there was not a signfiicant correlation between age and looking ($r(290) = .08, p = .15$). 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 ($\beta_{age} = .08$, $t = 19.34$, $p < .001$). 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 on two tasks: 1. Process social information to disambiguate naming events and discover the speaker's intended referent, and 2. Recall the learned mapping and use it to find the correct referent. The two kinds of naming events in this experiment were designed to vary in how much demand they placed on (1), but not on (2). To provide an additional test of our prediction, we fit a mixed effects model predicting children's looking on novel test trials from their age, their familiar word processing, and the naming event type, as well as the interaction of familiar processing and test type. This model showed a significant effect of age ($\beta_{age} = .05$, $t = 4.83$, $p < .001$), 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 (19, 30). 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 (31, 6). Children also acquire language at an impressive rate, for instance learning 4000–5000 words (11) by the time they are five. 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 (e.g., FIXME).

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 learn words as well as adults.

This social information processing bottleneck also characterized the learning of our sample of Autistic children. 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—Autistic children 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.

Materials and Methods

Participants. Data from typically-developing children was 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. 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 development 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 2, 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 development 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 Autistic children was collected at Stanford University, where parents and their children FIXME. 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.

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 labelled 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 32). 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 teach 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, the 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 these 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 it's 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 Experiment 1s and 2. Eight paired the named objects from the learning trials against their foils from the learning trials, four each 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 32), 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 Experiment 1, and 4.5s for Experiment 2. 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 2, learning trials were designed to separate demands on social attention across rather than within trials. In Experiment 2, we computed proportion of looking of the entirety of learning trials.

Due to occasional bouts of inattention, eye-gaze data was 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 was 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 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`.

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, and grants from the John Merck Scholars program, the Stanford Child Health Research Initiative to MCF .

References

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. *Psychological Science* 9(3):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. *Psychological Science* 17(8):684–691.
6. Bergelson E, Swingley D (2012) At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* 109(9):3253–3258.
7. Sokol S (1978) Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Research* 18(1):33–39.
8. Banks MS (1980) The development of visual accommodation during early infancy. *Child Development* pp. 646–666.
9. Forssberg H, Eliasson A, Kinoshita H, Johansson R, Westling G (1991) Development of human precision grip i: basic coordination of force. *Experimental Brain Research* 85(2):451–457.
10. Thelen E (1995) Motor development: A new synthesis. *American psychologist* 50(2):79.
11. Goulden R, Nation P, Read J (1990) How large can a receptive vocabulary be? *Applied Linguistics* 11(4):341–363.
12. Bates E et al. (1994) Developmental and stylistic variation in the composition of early vocabulary. *Journal of child language* 21(01):85–123.
13. Carey S, Bartlett E (1978) Acquiring a single new word. *Papers and Reports on Child Language Development* 15:17–29.
14. Yu C, Smith LB (2007) Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18:414–420.
15. Frank MC, Goodman N, Tenenbaum J (2009) Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* 20:578–585.
16. Scaife M, Bruner JS (1975) The capacity for joint visual attention in the infant. *Nature*.
17. Baldwin DA (1993) Early referential understanding : Infants' ability to recognize referential acts for what they are. *Developmental Psychology* 29:832–843.
18. Hollich GJ, Hirsh-Pasek K, Golinkoff RM (2000) Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monographs of the Society of Research in Child Development*.
19. Senju A, Csibra G, Johnson MH (2008) Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition* 108:303–19.
20. Corkum V, Moore C (1998) The origins of joint visual attention in infants. *Developmental psychology* 34(1):28.
21. Brooks R, Meltzoff AN (2005) The development of gaze following and its relation to language. *Developmental science* 8:535–43.
22. Csibra G, Gergely G (2009) Natural pedagogy. *Trends in cognitive sciences* 13(4):148–153.
23. 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(9):1443–1457.
24. Vida MD, Maurer D (2012) The development of fine-grained sensitivity to eye contact after 6years of age. *Journal of experimental child psychology* 112(2):243–256.
25. 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(11):e79659.
26. Dempster FN (1981) Memory span: Sources of individual and developmental differences. *Psychological Bulletin* 89(1):63.
27. Kail R (1991) Developmental change in speed of processing during childhood and adolescence. *Psychological bulletin* 109(3):490.

28. Gathercole SE, Pickering SJ, Ambridge B, Wearing H (2004) The structure of working memory from 4 to 15 years of age. *Developmental psychology* 40(2):177.
29. 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.
30. Vouloumanos A, Martin A, Onishi KH (2014) Do 6-month-olds understand that speech can communicate? *Developmental science*.
31. Thiessen ED, Saffran JR (2003) When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* 39(4):706–716.
32. Frank MC, Vul E, Saxe R (2012) Measuring the development of social attention using free-viewing. *Infancy* 17(4):355–375.