Salience and Social Cues in Early Word Learning

Daniel Yurovsky
yurovsky@stanford.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

Children learn their first words from social partners, but it is unclear to what extend they are attuned to these partners' social cues. Some theories argue that word learning is fundamentally social from its outset, with even the youngest infants understanding intentions and using them to infer a social partner's target of reference. In contrast, other theories argue that early word learning is fundamentally a perceptual process in which young children map words onto salient objects. Only gradually do children learn the predictive power of social cues like eye gaze weigh them more than perceptual cues like visual salience. We present a set of experiments that manipulate social and salience cues to reference across development. Our results show that children gradually improve in following social cues well into their fourth year, but that this does not produce decreased attention to salience. Further, we show that social and salience cues direct attention at different times, with social cues guiding attention at all ages during learning, but salience cues playing a role at test. Together, these results suggest that the debate may be fundamentally ill-posed, and suggest a new framework for thinking about the role of salience and social cues in early word learning.

Keywords: Language acquisition, word learning, attention, social cues, development

Introduction

How do young children learn the meanings of their first words? For example, when an adult produces a novel label in a complex natural scene, how can a child determine to which object—if any—the label refers (Quine, 1960; Bloom, 2000)? For adults, this problem is straightforward; in addition to learning a language, adults have learned to consult a speakers social gestures and use their understanding of a speakers communicative goals (Clark, Schreuder, & Buttrick, 1983). Social inference also characterizes the word-learning strategies of children late in their second year (e.g, Baldwin, 1991; Brandone, Pence, Golinkoff, & Hirsh-Pasek, 2007; Grassmann & Tomasello, 2010). But word learning likely begins much earlier, perhaps as early as at 6-months (Tincoff & Jusczyk, 1999; Bergelson & Swingley, 2012). Do very young children use social information to reduce referential uncertainty in early word learning?

Infants are situated in a social system from their first day of life. Some theories argue that infants leverage this social information from the very outset of word learning (Bloom & Markson, 1998; Waxman & Gelman, 2009). For instance, infants follow direction of gaze by 6-months (D'Entremont, Hains, & Muir, 1997) and are more likely to do so in the presence of other communicative signals (Senju, Csibra, & Johnson, 2008). Further, childrens successes in following gaze predict language development. Even more impressively, infants show some evidence of representing others beliefs, and these representations may affect their expectations by 7-months of age (Kovács, Téglás, & Endress, 2010). Infants may thus become sensitive to social cues through prelinguistic experience, and could, in principle, already use these cues from the outset of word learning (Bruner, 1983).

Nevertheless, competing theories argue that early word learning is primarily a perceptual process (Vygotsky, 1978). On these accounts, infants learn words by mapping them onto perceptually salient objects in their learning environments (Smith, 2000). Early child-directed naming events are characterized by multi-modal synchrony: mothers move the objects they label in temporal synchrony with the labels they speak (Gogate, Bahrick, & Watson, 2000), and the degree of synchrony predicts word-object mapping in young infants (Gogate, Bolzani, & Betancourt, 2006). Further, in studies that pit perceptual salience against social information (e.g., a speakers gaze), 10-month-old infants show no evidence of attending to gaze (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). Although 10-month-old infants may be able to follow gaze, they seem to treat it as less important than object salience in mapping words to objects. Further, when salience and gaze conflict, providing contradictory cues to the identity of the intended referent, 12- and 15-mo olds fail to learn any mappings at all (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Houston-Price, Plunkett, & Duffy, 2006). By 19- and 24- months, however, toddlers robustly learn labels for objects cued by gaze even in the presence of salient competitors (Moore, Angelopoulos, & Bennett, 1999; Hollich et al., 2000). These findings suggest a developmental trajectory in which infants learn to learn, gradually building skill in using social cues and discounting salience to learn the meanings of words.

**** The problem with these Hollich experiments is that you can't tell what's developing – social stuff, attentional stuff, or both. We set up two experiments that let us compare the coincident/conflict trials against gaze alone. We present first experiment 1 which is gaze alone, and then compare these results to those from Experiment 2 which replicates the Hollich design.

*** In addition, we can look at the effects of these cues on both learning and test. Preview result here about social cues winning at all ages in learning trials, but salience cues being important at test.

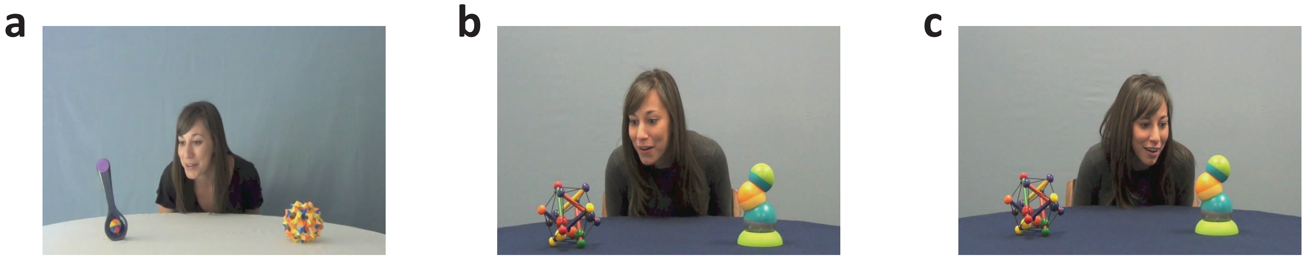

Figure 1: Example learning trials from Experiments 1 and 2. In Experiment 1 (a), the speaker turned towards one of the equally-salient toys and labeled it four times over the course of approximately 10 seconds. In Experiment 2, the speaker produced the same social cues and the same label as in Experiment 1, but the target object was either the more perceptually salient toy (b), or the less perceptually salient toy (c). Across experiments, we were thus able to determine the contributions of both salience and social information to early word learning outcomes.

Experiment 1

Method

*** SAY something about how we normed the stimuli here. Want to completely remove salience as a cue

Children's eye movements were tracked while they watched a series of naturalistic word-learning videos. In each, children saw a speaker seated at a table between two novel toys. She greeted them, and then turned towards one of the toys and labeled it three times in a short monologue. After these learning trial, children were tested for their knowledge of the referent for the new word using the Looking While Listening procedure (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). In addition, similar test trials were administered for known objects to measure children's processing of familiar words.

Participants Parents and their 1–4 year-old children were invited to participate in a short language learning study during their visit to the San Jose Children's Discovery museum. In total, we collected demographic and experimental data from 269 children, 122 of whom were excluded for one or more of the following reasons: abnormal developmental issues ($N = 27$), failure to calibrate ($N = 58$), and less than 75% exposure to English ($N = 36$). The final sample consisted of 27 1-1.5 year olds (9 girls), 19 1.5-2 year olds (7 girls), 38 2-2.5 year olds (13 girls), 26 2.5-3 year olds (10 girls), 15 4-3.5 year olds (9 girls), and 22 3.5-4 year olds (11 girls).

Stimuli The experiment consisted of two kinds of trials: learning and test. Learning trials were 12 second video clips in which a speaker first greeted the the child, and then turned towards one of the two toys on the screen, labeling it three times in a short monologue (Figure 4a). On the first learning trial, for example, the speaker said "Hi there! It's a *modi*. Look at the *modi*. What a nice *modi*."

Test trials followed the standard Looking While Listening protocol (Fernald et al., 1998). On each test trial, children saw two objects – one on each side of the screen – and heard a short audio clip of the speaker from the learning trials asking them to find a target object. Each test trials was 7 seconds long, and the target label was heard at 2.75 seconds. On *Familiar* test trials, both the target and distractor were common objects familiar to young children (e.g. book vs. dog). On *Novel* and *ME* test trials, children saw both of the toys from the previous learning trials, and were asked to find either the previously named toy (*modi*), or were asked about a novel label (*dax*).

In addition, the experiment contained two calibration checks: short videos in which small dancing stars appeared in four places on the screen. Because eye-tracker calibration can be imprecise, especially with younger children (Morgante, Zolfaghari, & Johnson, 2011), this check allowed us to adjust initial calibration settings to minimize the discrepancy between the behavior children produced and the behavior we analyzed (for details, see Frank, Vul, & Saxe, 2012).

Design and Procedure The experiment began with a 4-point calibration and then proceeded into a series of learning/test blocks. In each block, children first watched a learning trial in which a speaker labeled one of two on-screen toys. Following this learning trial, children were given a Looking While Listening test trial in which they saw both of these toys and were asked to find the toy labeled on the previous learning trial (e.g. "Can you find the *modi*?"). The entire experiment consisted of four learning trials, eight *Familiar*, six *Novel* test trials, and six *Mutual Exclusivity (ME)* test trials.

** Talk about Novel vs. ME trials as a salience/preference check. Foreshadow E2 results.

Data Analysis Children's eye movements during both learning and testing were analyzed using a Regions of Interest (ROI) approach. On learning trials, bounding-box ROIs were drawn by a human coder frame-by-frame for the speaker's face and for the two objects. On test trials, a bound-box ROI was drawn for each of the two static images. To ensure that recorded eye movements were mapped to the correct ROIs, children's calibrations were first adjusted by fitting a robust linear regression for their fixations during the calibration check video and using this model to transform eye movements during the rest of the experiment (Frank et al., 2012).

Children's learning and test behaviors were quantified by measuring their proportion of looking to each ROI on each trial. To ensure that proportions were representative, individual test trials were excluded from analysis if eye gaze data was missing for more than half of their duration. To compute age-group looking proportions, proportions were computed first for each individual trial, averaged at the individual-child level, and then averaged across children.

Window-of-analysis selection began by coding the point of disambiguation for each trial. This was the onset of the target label for test trials, and the rotation of the speaker's head for learning trials (marked '0' in the graphs in the Results section). The window for each trial began 1s after this point of disambiguation to allow children of all ages enough time to process and continued out to 3s after this point on both learning and test trials. To quantify children's learning with standard analyses, we aggregated these patterns of looking over time to compute the aggregate proportion of looking at the target object on each test trial.

Results

First, evidence from two analyses suggest that children were attentive to social cues during learning at all ages measured. First, for all age groups, looks to both target and distractor made up the minority of children's dwell times. Instead, children in all age groups spent more than 50% of their time attending to the speaker's face (Figure 2). Second, children were successful at attending to and following the speaker's social gaze even from the youngest ages measured. Children of all ages spent more time looking at the target than at the distractor during learning trials (smallest $t_1(23) = 3.20$, $p < .01$).

Analyses of test trials showed broad success on Familiar, Novel, and ME trials across development. The 1-year-olds trended towards significance on familiar trials ($t_1(26) = 1.65$, $p = .11$), and were non-significantly in the correct direction on Novel and ME trials. At all other ages, children looked to the target at above-chance levels on all test trials (smallest $t_{1.5}(17) = 2.10$, $p = .05$).

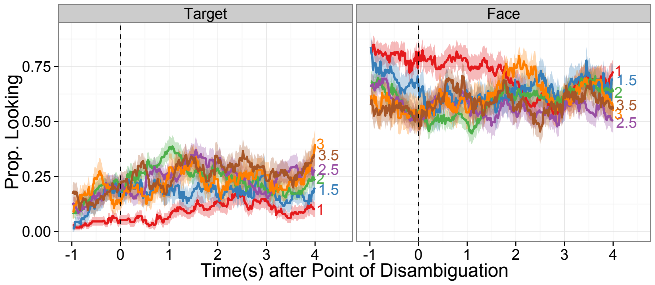

Figure 2: Children improved in their abilities to recognize familiar words, and to learn from both the *Extended* and *Brief* Cues over the course of development. Individual lines indicate different age groups and error bars indicate ±1SE.

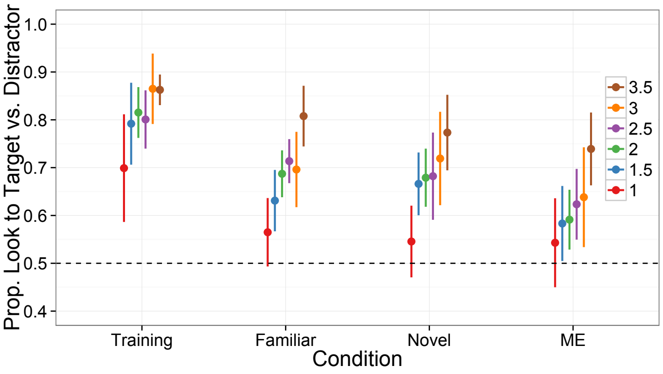

Figure 3: Proportion of time children fixated the correct correct target on each type of test trial in Experiment 1. Each dot indicates one half-year age group and each line represents a 95% confidence interval computed by non-parametric bootstrap. A proportion of .5 indicates chance performance.

In addition, children's abilities both to follow social cues during learning trials and to find the correct target on test trials improved across development. To quantify this improvement, we fit a mixed effects logistic regression to the data (Jaeger, 2008). This analysis revealed significant improvement across age ($\beta = .61$, $z = 4.03$, $p < .001$), as well a significant significant effect of Learning as compared to Novel trials ($\beta = 1.18$, $z = 3.11$, $p < .01$). No other effects or interactions approached significance. Figure 3 shows proportion of looking all kinds of trials at all ages.

Discussion

Together, these results provide clear evidence of a developmental trajectory in both word learning and word recognition (Figure 3).

***Improvements in following social gaze, but if it's cue weighting it seems like children are learning better weights really late into word learning. Sort of seems unlikely. Want to prime the intuition pump here about the importance of timing in learning and test.

Experiment 2

Method

Experiment 2 was identical to Experiment 1 in all respects except for the identity of the novel toys that served as the target and distractor. In contrast to Experiment 1, in which the two toys were balanced in their visual interest, the two toys in Experiment 2 were intentionally mismatched. For children in the *Salient* condition, the target was the more interesting toy, and the distractor the less interesting toy. In the *Non-Salient* condition, the identities of the toys were switched and the target was the less interesting toy. Thus, Experiment 2 allows us to investigate children's use of social cues to learn new words when they are aligned with salience, and when they are in opposition.

Participants Participants were recruited from the floor of the San Jose Children's Discovery museum as in Experiment 1. This time, however, we focused on the three youngest age groups. In the Salience condition, demographic and experimental data were collected from 117 children, 52 of whom were excluded for one or more of the following reasons: abnormal developmental issues ($N = 13$), failure to calibrate ($N = 25$), less than 75% exposure to English ($N = 33$), and inattentiveness ($N = 2$). The final sample consisted of 22 1-1.5 year olds (11 girls), 21 1.5-2 year olds (10 girls), 19 2-2.5 year olds (9 girls). In the Non-Salience condition, data were collected from 126 children, 71 of whom were excluded for one or more of the following reasons: abnormal developmental issues ($N = 9$), failure to calibrate ($N = 26$), and less than 75% exposure to English ($N = 36$). The final sample consisted of 26 1-1.5 year olds (13 girls), 25 1.5-2 year olds (11 girls), 15 2-2.5 year olds (4 girls).

Stimuli, Design, and Procedure Experimental stimuli were identical to those in Experiment 1, except that the identities of the novel toys were changed and new videos were recorded. In addition, Novel and ME test trials were updated to reflect the novel objects used in Experiment 2. The procedure, including the order of the trials, was identical.

Results

***No change in training trials!!! Salience is getting massively overwhelmed by the face, and by gaze-following. ***Huge effect on test trials, but age-independent. Kids definitely don't seem to be *unlearning* salience weights

Table 1: Predicting Learning of Novel Words.

Predictor	Estimate (SE)	t-value	Significance
Intercept	-.15 (.48)	-.31	$p = .75$
Age (yrs)	.71 (.23)	3.11	$p < .01$
Salient	.86 (.47)	1.82	$p = .07$
NonSalient	-.89 (.37)	-2.40	$p < .05$
ME	-.46 (.36)	-1.27	$p = .20$
Familiar	.23 (.38)	.596	$p = .55$
Learining	1.08 (.46)	2.35	$p < .05$
Sal*ME	-2.01 (.60)	-3.37	$p < .001$
NonSal*ME	1.67 (.54)	3.07	$p < .01$
Salient*Fam	-.49 (.65)	-.76	$p = .45$
NonSal*Fam	1.35 (.57)	2.37	$p < .05$
Salient*Learn	-.15 (.83)	-1.85	$p = .85$
NonSal*Learn	.77 (.65)	1.20	$p = .23$

Discussion

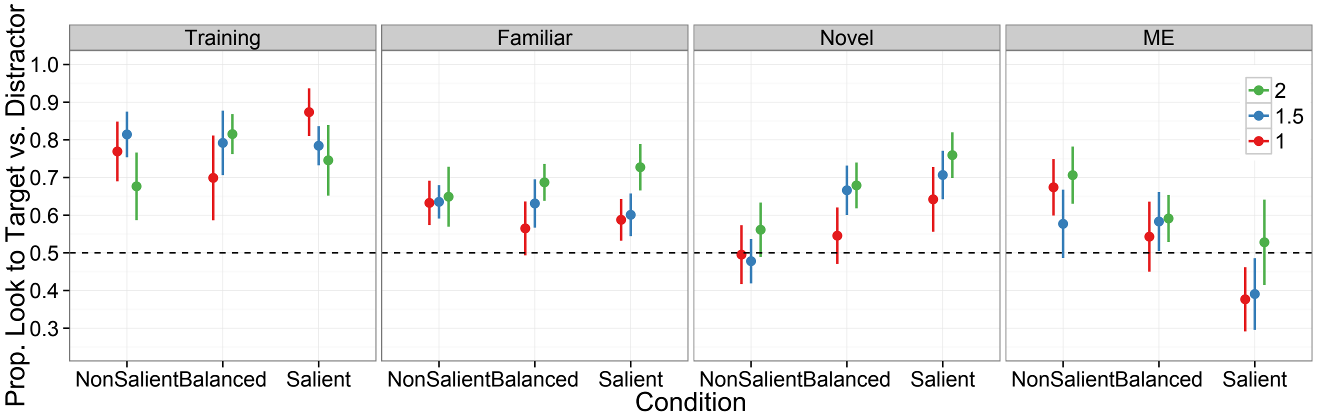

Figure 4: .

Conclusion

Main ideas: face is relevant, attended to, and used to find the target. The reason why could change over development (e.g. face itself could be salient), but either way "social cues" get used.

Salience definitely matters, but mostly at test. Could be a cue-weighting model in which once the face is gone the salience cue gets high weight. But something about cue weighting seems wrong based on learning trial behavior. Unclear how to predict developmental differences unless cues are not normalized. Also, in general, this kind of model really misses the inherent temporal aspects of word learning.

Salience could still be used in the absence of social information for either smart or dumb reasons, and might end up being adaptive. But it seems pretty clear that we're not seeing a re-weighting across development of these two cues.

Seems like what we really want to pay attention to are memory, attentional control, and timing. Probably all of these are developing?

Maybe the two timescales idea (Frank, Goodman, & Tenenbaum, 2009; McMurray, Horst, & Samuelson, 2012) is the right way to think about this kind of thing. Performing above competence?

Acknowledgments

We are grateful to Janelle Klaas for collecting the data, and to all of the members of the Language and Cognition Lab for their feedback on this project. In addition, we thank the parents, children, and staff at the San Jose Children's Discovery Museum for supporting us in collecting developmental data. This work was supported by a NIH NRSA F32HD075577 to DY as well as grants from the Merck Scholars Foundation and the Stanford Center Health Research Initiative to MCF.

References

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875–890.

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*, 3253–3258.

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

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*, *2*, 67–73.

Brandone, A. C., Pence, K. L., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Action Speaks Louder Than Words: Young Children Differentially Weight Perceptual, Social, and Linguistic Cues to Learn Verbs. *Child Development*, *78*, 1322 – 1342.

Bruner, J. (1983). *Childs talk: Learning to use language*. Oxford: Oxford University Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*, 245–258.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A Demonstration of Gaze Following in 3- to 6-Month-Olds. *Infant Behavior and Development*, *20*, 569–572.

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

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

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

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878–94.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to Maternal Multimodal Naming by 6- to 8-Month-Old Infants and Learning of Word-Object Relations. *Infancy*, *9*, 259–289.

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*, 252–263.

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

Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of experimental child psychology*, *95*, 27–55.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831–877.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental psychology*, *35*, 60–8.

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2011). A Critical Test of Temporal and Spatial Accuracy of the Tobii T60XL Eye Tracker. *Infancy*, 1–24.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, *77*, 266–280.

Quine, W. V. O. (1960). *Word and Object* (Vol. 22).

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition*, *108*, 303–19.

Smith, L. B. (2000). How to learn words: An associative crane. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the Word Learning Barrier* (pp. 51–80). Oxford, UK: Oxford University Press.

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, *10*, 172–175.

Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Science*, *13*, 258–263.