Beyond Naïve Cue Combination: Salience and Social Cues in Early Word Learning

Daniel Yurovsky and Michael C. Frank

Department of Psychology, Stanford University

Author Note

Please address correspondence to:

Daniel Yurovsky

Jordan Hall (Building 420)

Stanford University

450 Serra Mall

Stanford, CA 94305

Email: yurovsky@stanford.edu

Abstract

Children learn their earliest words through social *interaction*, but it is unknown how much they rely on social *information*. 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 largely a perceptual process in which young children map words onto salient objects. One way of unifying these accounts is to model word learning as weighted cue-combination in which children attend to many potential cues to reference, but only gradually learn the correct weight to assign each cue. We tested 3 predictions of a naïve cue-combination account using an eye-tracking paradigm that combines social word-teaching and two-alternative forced-choice testing. None of the predictions were supported. Thus, while aspects of this unifying account are correct, it must be amended to capture the dynamics of children's behavior across differing referential situations. In addition, we discuss methodological implications for research using two-alternative displays.

Keywords: Language acquisition, word learning, attention, social cues, eye-tracking, cognitive development

Beyond Naïve Cue Combination: Salience and Social Cues in Early Word Learning

How do children learn the meanings of their first words? A number of influential theories conceptualize the primary learning mechanism available to young infants as one that makes association between perceptual stimuli (Piaget, 1952; Vygotsky, 1978). On these kinds of accounts accounts, infants learn the meanings of labels like "ball" and "dog" by mapping them onto salient objects in their learning environments (Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Smith, 2000). This kind of account is appealing because the mechanisms it requires for the onset of word learning—perceptual orienting and associative mapping—are universally agreed to be in the repertoire of young infants (e.g. Fantz, 1964; Haith, 1980; although c.f. Spelke, 1998). In addition, the ecological context of language learning appears to support perceptually-driven learning. For instance, 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 successful word-object mapping for young infants (Gogate, Bolzani, & Betancourt, 2006).

However, infants are situated in a social system from their first day of life. Thus alternative theories argue that infants may leverage social information when learning even their first words (Bruner, 1983; Bloom & Markson, 1998). 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, individual differences in children's gaze-following predict differences in vocabulary development (Brooks & Meltzoff, 2008). In addition, in some experiments infants appear to be representing others' beliefs, and these representations affect their expectations by 12- or even 6-months of age (Vouloumanos, Onishi, & Pogue, 2012; Vouloumanos, Martin, & Onishi, in press). Infants are thus tuned to social cues, and could in principle already use these cues from the outset of word learning.

Because these two classes of theories are posed as mutually-exclusive competitors,

and because both are supported by compelling empirical evidence, each has attempted to re-conceptualize the evidence in favor of the other on its own terms. For example, researchers in the perceptual tradition have shown that a cases of putatively 'social' understanding can be explained as a set of learned perceptual associations (e.g. Goldstein & Schwade, 2008; Yu & Smith, 2012a; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014). On the other side, researchers from the social tradition have argued that the perceptual signals shown to drive learning are effective precisely because infants infer that they are being presented by a social, pedagogically motivated caregiver (Csibra, 2010; Deligianni, Senju, Gergely, & Csibra, 2011).

In part due to this stalemate, a third alternative has been to try to unify these two accounts under a single framework. One possible unification is a model in which infants are sensitive to many cues to reference: both perceptual cues like visual salience and temporal contiguity *and* social cues like eye-gaze and pointing. To determine the referent of a speaker's utterance, children could *combine* all of the available cues, assigning each a weight proportional to its predictive validity. On such an account, developmental changes in determining a speaker's target of reference are due to a process of learning the correct weights to assign to each kind of cue. Early on, children may be biased to assign high weight to perceptual cues. However, over development, children might gradually assign higher weight to social cues as they learn that they are better predictors of reference (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Golinkoff & Hirsh-Pasek, 2006).

Developmental Cue Combination

Support for a developmental cue-combination account comes from studies that pit perceptual salience against social information (e.g., speaker gaze) at different developmental ages. Hollich et al. (2000) presented 12 studies which in which they varied the referential cues highlighting two different objects in an ambiguous naming events. In the first three experiments, one of the objects was perceptually salient, and one was fixated

by the speaker. Analyses compared a condition in which the same object received both cues (Coincidental) to a condition in which the cues pointed to different objects (Conflict). In this study, 19- and 24-month-olds appeared to assign more weight to the social cue, preferentially mapping the label onto the object the speaker fixated regardless of which was more salient. In contrast, although 12-month olds showed some evidence of following the speaker's gaze in training, they looked more at the more salient object at test in both conditions. A followup experiment from Pruden, Hirsh-Pasek, Golinkoff, and Hennon (2006) showed that 10-month olds did not attend to the speaker's gaze at all.

However, in almost all of these studies, the target and competitor objects remained in the same position during both training and test. Thus, it is unclear in most of this data whether infants mapped the label onto an object or onto a location (e.g., Benitez & Smith, 2012). In the two experiments in which target position switches from training to test, learning is disrupted in all of the age-groups except for the 24-month-olds (Hollich et al., 2000; Pruden et al., 2006). This is consonant with earlier data from Moore, Angelopoulos, and Bennett (1999) who showed in a head-turn procedure that conflicting perceptual and social cues disrupted learning in 18- but not 24-month-olds.

Together, these studies present an intriguing but incomplete set of supporting data for a developmental cue combination account. First, these studies present an incomplete picture of children's behavior during training trials. Because these studies did not track children's eye-gaze, they do not capture the dynamics of attention in the presence of different cues to reference. That is, they characterize children's preference for one object over the other, but not how children allocated their attention to the speaker and toys over time. Second, it is unclear from this data whether a change in the relative strengths of social and perceptual cues is due to an increase in weight for social cues, a decrease in perceptual cues, or both. Answering this question would require comparing the two conditions in which cues are an opposition to a condition in which *only* the social cue is available. While Moore et al. (1999) did run this condition, they have relatively low power

to detect differences between them due to small samples and a challenging test: forced choice responding. Hollich et al. (2000) ran a number of gaze-alone with 12-month-olds, but not with the older age groups.

The goal of the studies in this paper was to gather a larger, developmentally broader set of eye-tracking data that would allow us to answer these questions.

** WHERE DOES THIS SECTION GO???**

Of course, children must learn more than mappings between labels and objects in the world. While object labels represent a large slice of typical early vocabularies (Caselli et al., 1995; Tardif, Fletcher, Liang, Zhang, & Kaciroti, 2008), infants also learn verbs, adjectives, and interjections (Fenson et al., 1994; Clark, 2003; Bergelson & Swingley, 2013). It may turn out that the kinds of mechanisms advanced in perceptual accounts of early word learning are insufficient to explain this diversity of word meanings (e.g. Bloom, 2000; Waxman & Gelman, 2009). However, while this remains an open question (e.g. Smith, 2000; Piantadosi, Tenenbaum, & Goodman, 2012; Scott & Fischer, 2012), we take concrete nouns to be an important case study of the roles of perceptual and social information in early word learning[1].

Predictions of Developmental Cue Combination

Weighted cue-combination is an intuitive, computationally simple model of the process of change in early word learning. For example, a number of computational models have implemented a version of this idea (Frank, Goodman, & Tenenbaum, 2007; Frank, Tenenbaum, & Fernald, 2013). It is also consistent with properties of our perceptual system: Within and across modalities, adults weigh cues in proportion to their predictive power, combining them as predicted by ideal observer models (Ernst & Banks, 2002; Jacobs, 2002). Yet a number of its detailed predictions remain untested.

[1]For convenience we refer to "word-object mapping" and "word learning" interchangeably. But, of it is only part of the process of learning a phonological form, mapping it to a referent, retaining this mapping, and determining how it generalizes.

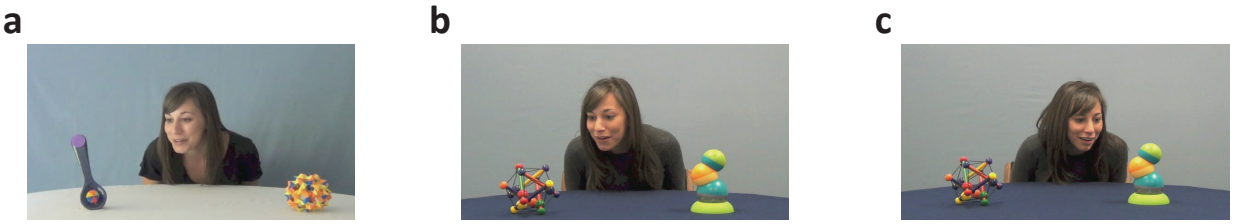

Figure 1. Example learning trials from Experiments 1 and 2. In Experiment 1 (a), the speaker turned towards one of the equally-salient toys and labeled it three times over the course of ∼10 seconds. In Experiment 2, the speaker produced the same social cues and the same label as in Experiment 1, but the target object was either the more perceptually salient toy (b), or the less perceptually salient toy (c). These manipulations allowed us to measure the contributions of both salience and social information to word-object mapping.

Using eye-tracking to measure how children learn word-referent mappings in ambiguous naming events, we test three predictions of the cue-combination model of developmental change:

1. Developmental change is due to re-weighting across cues,
2. Cue weights drive attention during learning, and
3. Perceptual cues decrease in weight across early development.

In two experiments, we show that none of these predictions is supported. Thus, while cue-combination captures important insights about early word learning, a naïve version of this account is insufficient to explain the observed developmental trajectory.

Experiment 1

In almost all previous experiments investing cue-combination in early word learning,social cues were always pitted against perceptual cues (c.f. Moore et al., 1999). Thus, their results indicating developing preferences for social information over perceptual information are consistent with three possible explanations: (1) Social cues increase in weight, (2) Perceptual cues decrease in weight, and (3) Perceptual and Social Cues both

change in their *relative* weight. Experiment 1 was designed to distinguish between these three possibilities by measuring the development of children's abilities to follow and learn from social gaze in the *absence of competing salience cues*. A naïve cue-combination account, in which developmental changes in cue use result from learning their relative predictive weights, makes a null prediction: children's responses should not change significantly across development when only one cue is available.

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 trials, children were tested for their knowledge of the referent for the new word using the preferential looking procedure. In addition, to measure children's processing abilities for familiar words, similar test trials were administered with known items.

Method

Stimulus Norming. To minimize salience differences between the two potential referents, we first normed them using aggregate adult judgments. Thirty-eight adults on Amazon Mechanical Turk were shown toys two at a time from a set of 10. For each pair, they were asked to pick the toy they would rather play with. Each participant made 20 choices, with toys sampled at random, producing $\sim$7.6 responses for each pair of toys. Based on these responses, we selected the two toys that were best balanced against each other (see Figure 2a).

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

($N = 36$).[2] 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 and Design. The experiment consisted of two kinds of trials designed to measure both how children allocate their attention while learning from a social partner, and what word-object mapping information they extract from these learning events. Learning trials were ∼12s 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 2a). On the first learning trial, for example, the speaker said "Hi there! It's a *modi*. Look at the *modi*. What a nice *modi*."

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 trial was 7s long, and the target label was heard at 2.75s. On *Familiar* test trials, both the target and competitor were common objects familiar to young children (e.g. book vs. dog). On *Novel* and *Mutual Exclusivity (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*). These ME trials were designed as a strong test of mapping formation; looking to the correct target on Novel trials could result from familiarity or preference rather than mapping. However, correct performance on both Novel and ME trials could only result from knowledge of the specific label used in training.

Finally, the experiment contained two calibration checks: short videos in which small dancing stars appeared in four places on the screen. These checks allowed us to adjust initial calibration settings when they were imprecise (for details, see Frank, Vul, & Saxe, 2012).

[2]These exclusion criteria were preset in this study on the basis of previous work (Yurovsky, Wade, & Frank, 2013).

Procedure. The eye-tracker was first calibrated for each child using a 2-point calibration. Next, children saw four learning trials in which the speaker looked at one of two toys on the screen and labeled it three times. Finally, children saw all of the test trials, in which their knowledge of both familiar and novel word-object mappings was tested. Two calibration checks (described above) were embedded in the learning phase. The entire experiment consisted of 4 learning trials, 8 Familiar, 6 Novel, and 6 ME test trials.

Data Analysis. Children's eye movements during both learning and testing were analyzed using a Regions of Interest (ROI) approach. Bounding-box ROIs were drawn by a human coder for the speaker's face (learning trials) and for the two objects (learning and test trials). Children's calibrations were adjusted by fitting a robust linear regression for their fixations relative to known locations on calibration check videos. These regressions were used to transform eye movements for all learning and test trials (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 were 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. 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 learning with standard analyses, we aggregated these patterns of looking over time to compute proportion of target looking on each test trial. JUSTIFY WINDOW CHOICE?

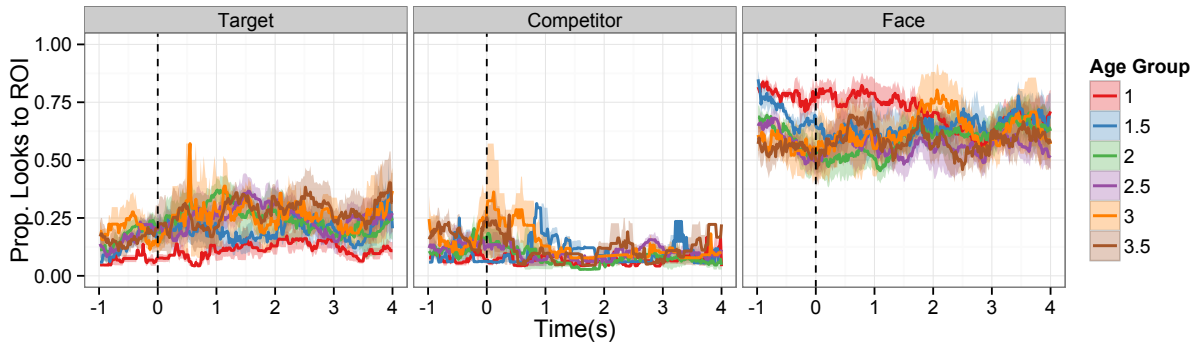

Figure 2. Looking behavior during training trials in Experiment 1. Children in all age groups spent the majority of training trials fixating the speaker's face.

Results

I THINK WE NEED TO DISCUSS THE RESULTS PER SE BEFORE TESTING PREDICTIONS IN THIS LONGER FORM PAPER.

E.G. SHOW TRAINING ROI CURVES THAT WE CUT FROM THE COGSCI PAPER?

ALSO RECOGNITION CURVES FOR FAMILIAR, NOVEL, ME - SO CAN BE COMPARED WITH PRIOR WORK AND SO WE CAN SEE WHAT'S GOING ON WITH YOUNGEST KIDS.

In Experiment 1, we address two predictions of naïve cue combination: how cues affect attention during learning, and how weights change across development.

Older children were better at *disengaging* from social stimul. 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 competitor during learning trials (smallest $t(23) = 3.20$, $p < .01$; Figure 4). However, for all age groups, looks to both target and competitor 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 3).[3] Thus, the primary driver of developmental change was

[3]All data and code for analysis available at `http://github.com/dyurovsky/ATT-WORD`.

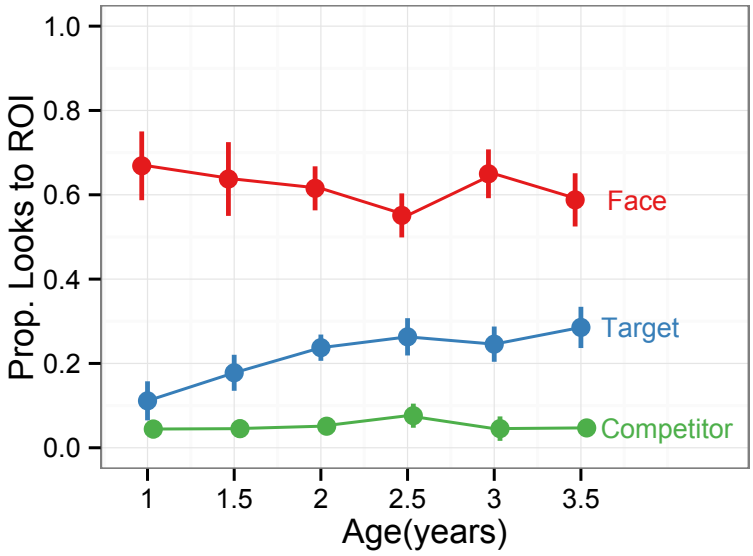

Figure 3. Proportion of children's looking the target toy, competitor toy, and the speaker's face during learning in Experiment 1. Children of all ages spent the majority of the learning trials looking at the speaker's face. Disengaging from the face and fixating the target increased across development. Error bars indicate 95% confidence interval computed by non-parametric bootstrap.

not stronger discrimination between the target and competitor (predicted by greater social cue weights), but rather improved ability to disengage from the speaker's face.

Developmental change was not primarily due to re-weighting. In line with the naïve cue combination account, attention due to the social cue during learning carried forward to correct mapping at test. Analyses of test trials showed broad success on Familiar, Novel, and ME trials across development. The 1–1.5 year-olds trended towards significance on familiar trials ($t(26) = 1.65$, $p = .11$), 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 $t(17) = 2.10$, $p = .05$).

However, 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

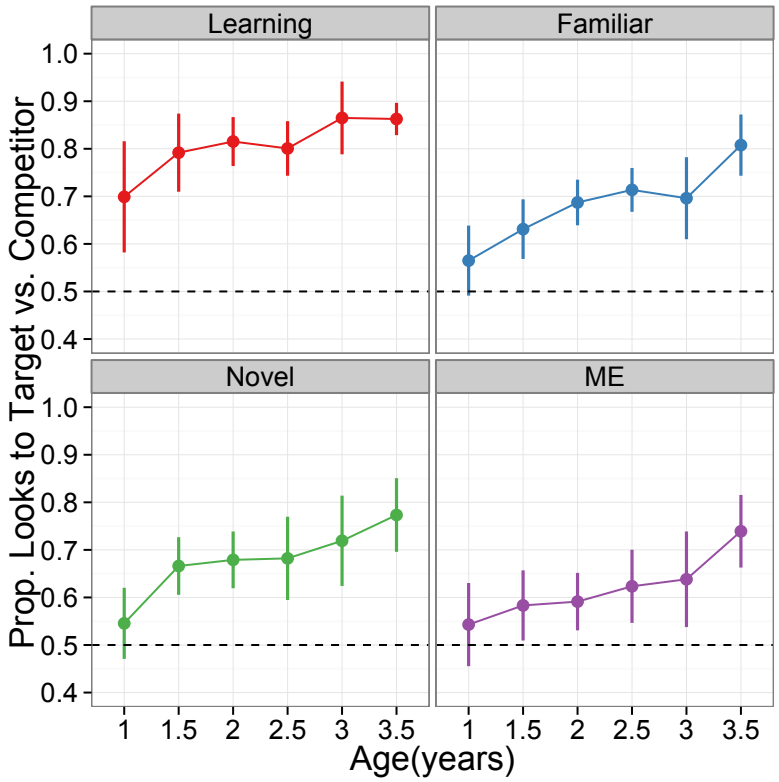

Figure 4. Proportion of time children fixated the correct target on each type of test trial in Experiment 1. Children improved on all measures across development. Each dot indicates one half-year age group and each line represents a 95% confidence interval computed by non-parametric bootstrap. A proportion of .5 indicates chance performance.

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 as 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 4 shows proportion of looking to the correct target for all kinds of trials at all ages.

Thus, across development, children improved in learning from the social cue, even when it was the only cue available. This suggests that relative re-weighting across cues is not the only driver of improved word learning.

Discussion

Together, these results provide evidence both of early competence in the use of social gaze to determine the target of a speaker's reference, as well as improvement across development. Further, improvements in gaze-following also paralleled improvements in both finding the referents of these novel words on subsequent test trials, and also finding the referents of familiar words (Figure 4).

These results thus provide support for one key claim of the developmental cue-combination account: children are sensitive to social cues quite early. Young children could assign small—but non-zero—weight to social cues, and then gradually assign them more weight over development. However, the results also provide evidence *against* the predictions that cues drive attention, and that developmental change is due to relative re-weighting. First, children of all ages found the speaker's face highly engaging, and spent the majority of their time fixating it rather than the referents on learning trials. The primary behavioral development was the ability to *disengage* from the speaker's face. Second, children showed gradual improvement in fixating the target during both learning and test trials well into their fourth year.

This data could be consistent with a modified version of the cue-combination account in which cues change in both their absolute and relative weights due to learning. However, while children undeniably encounter naming events in their third and fourth years, it seems unlikely that the process of learning the validity of social gaze would extend over such a long period of time.

TRANSITION?

In Experiment 2 we manipulated the relative salience of the target and competitor objects children learned about (c.f. Hollich et al., 2000). This allowed us to measure how salience affects children's looking during both learning and test, providing a test of all three predictions of the naïve cue-combination account.

Experiment 2

Experiment 2 was identical to Experiment 1 in all respects except for the identity of the novel toys that served as the target and competitor. In contrast to Experiment 1, in which the two toys were balanced in their visual salience, the two toys in Experiment 2 were mismatched. For children in the *Salient* condition, the target was the more interesting toy, and the competitor the less interesting toy. In the *Non-Salient* condition, the identities of the toys were switched—the target was the less salient toy. Experiment 2 allowed us to investigate children's use of social cues to learn new words both social cues and salience indicate the same referent, and when they are in competition (as in Hollich et al., 2000; Pruden et al., 2006).

Method

Participants. Participants were recruited from the floor of the San Jose Children's Discovery museum as in Experiment 1. For Experiment 2, we focused on the three youngest age groups. In the Salient 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-Salient 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. The procedure, including the order of the trials, was identical.

Results and Discussion

AGAIN, NEED BASIC DESCRIPTIVE ANALYSES: WERE PARTICIPANTS ABOVE CHANCE? WHAT DID THEY LOOK AT, ETC...

To determine the effect of perceptual salience on word learning, we compared children's looking in the Salient and Non-Salient conditions not only to each other, but also to the Balanced condition tested in Experiment 1.

Perceptual salience did not drive attention during learning. In contrast to the prediction of the naïve cue-combination account, children's looking behavior during learning trials was not significantly affected by the salience of the target and competitor (Figure 5, top). As in Experiment 1, children of all ages spent the more time looking at the target than the competitor, but looking time to both made up the minority of their dwell time; children spent the majority of learning trials looking at the speaker's face (smallest proportion—2-year-olds in the Non-Salient Condition: .51).

This null-result could be due to the toys being too similar in their salience, making this a weak test of the cue-combination model. However, salience exerted a strong effect on test trials—children in all age groups were strongly attracted to the salient object. When the target referent was salient, children at all ages looked at it for the majority of the window of analysis on Novel test trials (smallest $t(19) = 2.96$, $p < .01$). When the target was non-salient, no age group look showed evidence of learning on Novel test trials (largest $t(13) = 1.46$, $p = .17$). Mutual-exclusivity (ME) trials showed the opposite pattern. When the target referent was salient, children in the two younger age groups looked at the correct referent on ME trials (the competitor) at *below* chance levels (smallest $t(20) = -2.29$, $p < .05$). In the Non-Salient condition, even the youngest children looked at the correct referent on ME trials at above chance levels (smallest $t(22) = 4.51$, $p < .001$). Figure 5 (middle and bottom) shows looking behavior at test in both Experiments 1 and 2.

Perceptual cues did not decrease in weight across development. The effect of perceptual cues at test did not appear to change across the 1–2 year range. We fit a

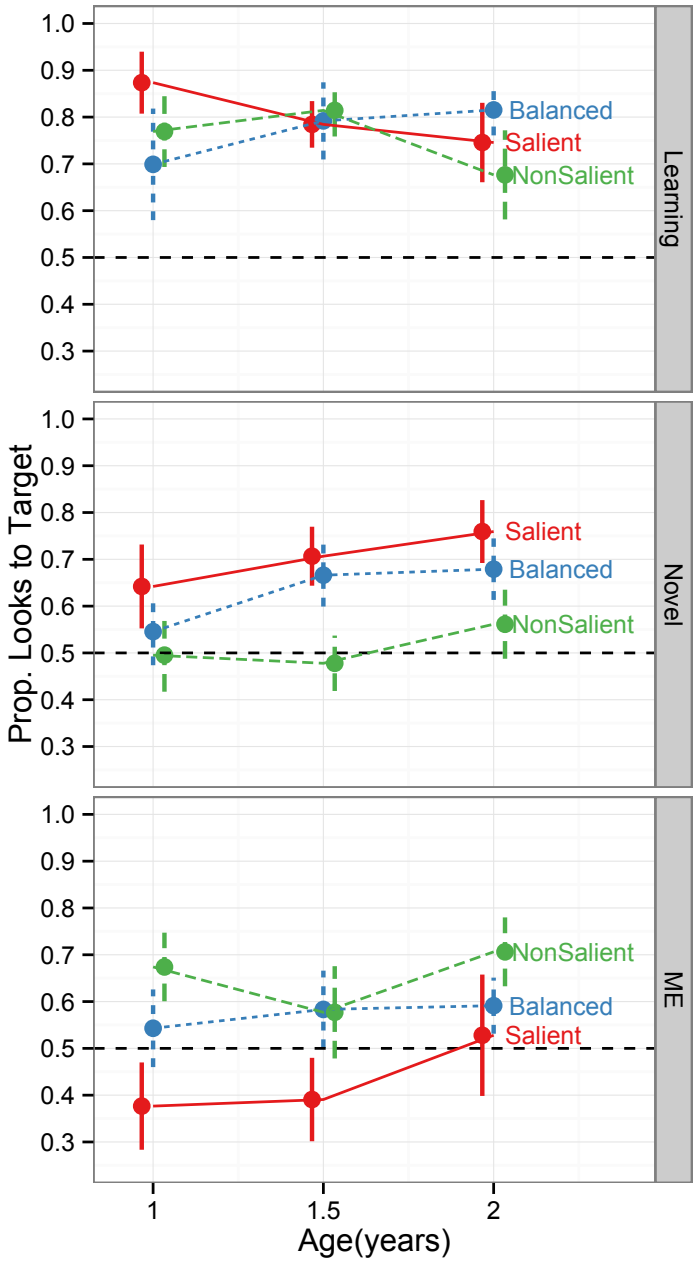

Figure 5. Proportion of time children fixated the correct target on Learning and Test trials in Experiments 1 (three youngest age groups) and 2. Salience had the predicted effect on looking behavior at test, but relatively little during learning. Each dot indicates one half-year age group and each error bar represents a 95% confidence interval computed by non-parametric bootstrap. A proportion of .5 indicates chance performance.

Table 1

Mixed-effects Logistic Regression Model for Looking Behavior in Experiments 1 and 2.

Predictor	Estimate	Std. Error	z value	p value	
Intercept	-0.63	0.63	-0.99	0.32	
Age(years)	0.43	0.27	1.61	0.11	
Familiar	1.53	0.73	2.10	0.04	*
Salient	0.92	0.48	1.90	0.06	.
NonSalient	-1.00	0.37	-2.70	0.01	**
Learning	0.94	0.44	2.11	0.03	*
ME	-0.32	0.36	-0.89	0.37	
Salient × Learning	0.00	0.84	0.00	1.00	
Non-Salient × Learning	1.15	0.65	1.76	0.08	.
Salient × ME	-2.23	0.61	-3.65	0.00	***
Non-Salient × ME	1.59	0.54	2.92	0.00	**

mixed-effects logistic regression to the data from both experiments to determine how age and experimental condition impacted looking behavior during both learning and test. After controlling for performance on Familiar trials, this regression showed a significant effect of condition, and an interaction between trial type and condition. Children looked more to the salient object at test regardless of whether it was the target or competitor, and significantly more at the target during learning trials regardless of whether it was salient. None of these factors interacted with age (Table 1).

Developmental change was not due to re-weighting across cues. Together with the t-tests above, the mixed-effects model suggests that children are not relatively re-weighing salience and social cues over the course of development. While salience certainly plays a role in directing looking behavior, it does not appear to play a role during learning itself. However, salience has a strong effect during test: In the absence of any

social information, salience directs children's attention in a way that does not appear to change over early development.

General Discussion

Is children's early word-object mapping fundamentally social, or is it mostly driven by perceptual processes? A weighted cue-combination account provides a simple framework to unify social and perceptual factors in early word learning (Hollich et al., 2000; Frank et al., 2013). Under this kind of account, perceptual cues are weighed higher in early learning, while social cues gradually gain weight as children learn their predictive power across early naming events. We tested this account in two word-learning experiments and found that its predictions were inconsistent with the data.

Although a naïve cue-combination account would predict that developmental change is largely driven by the relative re-weighting of cues, our data showed little evidence of this (contra prediction 1). Instead, developmental changes during learning appeared to be driven by disengagement from the social stimulus, not disengagement from the perceptually salient target (contra prediction 2). Finally, perceptual salience exerted its effects mostly at test, and did so consistently across early development instead of decreasing in weight (contra prediction 3).

These results present a picture of early word learning that is consistent with the spirit of cue-combination accounts like the Emergentist Coalition model, but which requires significant refinement of their details. For instance, although perceptually salient toys may attract attention and direct looking and learning in the absence of social partners, the speaker herself was the most interesting aspect of our naming events for young children. Although it is unclear to what extent young children's attention to the speaker reflects their understanding of her intentions, it is quite clear that their attention was both attracted and directed by social information. The primary change in cue-use across development may thus be due decreasing weight on perceptually salient toys, but rather to

developing inhibitory control more broadly.

NEW PARAGRAPH: ONLY OK. Our work here has significant implications for users of two-alternative preferential looking displays. We found that, when alternatives were not matched for perceptual salience, the relatively more salient object dominated children's looking preferences for all age groups. In particular, we saw evidence of novel word learning for the 1.5–2-year-olds in the balanced salience condition, but this result was masked if the target item was more salient and exaggerated if the target was less salient. Especially for these young participants, small differences in the perceptual properties of the stimuli may mask learning, presumably because overcoming perceptual salience requires inhibitory control that these young children do not have. Fernald, Zangl, Portillo, and Marchman (2008) discusses the utility of matching displays for salience in order to achieve precise measurements with young children. Our dat ratify this suggestion and suggest that such matching is especially important when comparing across ages, since children's changing inhibitory capacities would otherwise pose a confound in interpreting changes in looking.

COULD BE INTERESTING TO LOOK AT WHETHER THE AVERAGE OF SALIENT AND NON SALIENT IS DIFFERENT AT ALL FROM BALANCED?

Learning a new word relies on processes that work at multiple time-scales. Children need to identify a speaker's referent in-the-moment, encode a mapping between the label and referent, recall multiple labeling events and integrate acrobatss them, and use their learned mappings to identify the object in novel contexts (Frank, Goodman, & Tenenbaum, 2009; McMurray, Horst, & Samuelson, 2012; Yu & Smith, 2012b). We have provided data here that falsify some predictions of a naïve cue-combination model. But our critiques converge with a broader theoretical problem: naïve cue-combination does not distinguish among the component problems that word learners must solve. In our experiments, for instance, children used different cues to identify a speaker's referent and to find it in a novel test context. Building a more satisfying model of the development of word learning will require integrating the cues children use to identify referents with an understanding of

how these cues interact with attentional control, memory, and the conversational contexts in which naming occurs (Frank et al., 2013; Yurovsky et al., 2013).

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 NIH NRSA F32HD075577 to DY as well as grants from the Merck Scholars Foundation and the Stanford Center Health Research Initiative to MCF.

References

Benitez, V. L., & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*, 339-352.

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*, 391–397.

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.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220.

Bruner, J. (1983). *Child's talk*. Oxford: Oxford University Press.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159–199.

Clark, E. V. (2003). *First language acquisition*. Cambridge University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*, 141–168.

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*, 270–281.

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental psychology*, *47*, 1499–1503.

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*(4), 569–572.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a

statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, (5, Serial No. 242).

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

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., Goodman, N. D., & Tenenbaum, J. B. (2007). A bayesian framework for cross-situational word-learning. In *Nips.*

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24.

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.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science*, *19*, 515–523.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby Wordsmith: From Associationist to

Social Sophisticate. *Psychological Science*, *15*, 30–34.

Haith, M. M. (1980). *Rules that babies look by: The organization of newborn visual activity.* New Jersey: Lawrence Erlbaum Associates, Inc.

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.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*(8), 345–350.

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.

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*, 60–68.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International University Press.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*.

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.

Scott, R. M., & Fischer, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163–180.

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.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowlege. *Infant Behavior and Development*, *21*, 181–200.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., & Kaciroti, N. (2008). Baby's first 10 words. *Developmentla Psychology*, *44*, 929–938.

Vouloumanos, A., Martin, A., & Onishi, K. H. (in press). Do 6-month-olds understand that speech can communicate? *Developmental Science*.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences*, *109*(32), 12933–12937.

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.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*, 1289–1309.

Yu, C., & Smith, L. B. (2012a). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262.

Yu, C., & Smith, L. B. (2012b). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*, 21–39.

Yurovsky, D., Wade, A., & Frank, M. C. (2013). Online processing of speech and social information in early word learning. In M. Knauff, M. Pauen, N. Sebanz, &

I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1641–1646). Austin, TX: Cognitive Science Society.