



PAPER

Beyond naïve cue combination: salience and social cues in early word learning

Daniel Yurovsky and Michael C. Frank

Department of Psychology, Stanford University, USA

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 four predictions of this kind of 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. We thus propose an alternative unifying account: children are sensitive to social information early, but their ability to gather and deploy this information is constrained by domain-general cognitive processes. Developmental changes in children's use of social cues emerge not from learning the predictive power of social cues, but from the gradual development of attention, memory, and speed of information processing.

Research highlights

- Measured development of perceptual and social cue use in word learning
- Compared data to predictions of a naïve cue combination account
- Found evidence of early sensitivity to social cues
- Showed that developmental change was due to domain-general cognitive processes

Introduction

How do children learn the meanings of their first words? A number of influential theories conceptualize young infants' primary learning mechanism as making associations between perceptual stimuli (Piaget, 1952; Vygotsky, 1978). On these kinds of 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). These accounts are appealing on grounds of parsimony: The mechanisms they require 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).

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).¹ Thus, associations between perceptual

¹ For convenience here—and reflecting the orientation of our work towards concrete noun learning—we refer to 'word-object mapping' and 'word learning' interchangeably. Of course, mapping is only a small part of the process of learning a word, which includes at least: learning a phonological form, mapping this form to a referent, inferring the meaning that licenses this reference and how it generalizes across instances, and retaining all of this information in memory for future use.

stimuli is an attractive account for the mechanisms of early word learning.

Infants are situated in the social world from their first day of life, however, and this social world is the *source* of the multi-modal structures in their perceptual input. Alternative theories of word learning thus argue that infants leverage social—not just perceptual—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), especially in the presence of other communicative signals (Senju, Csibra & Johnson, 2008). Further, individual differences in gaze-following predict differences in vocabulary development (Brooks & Meltzoff, 2008). In addition, in some experiments infants appear to be representing others' communicative goals, and these representations affect their expectations about language by 12 or even 6 months of age (Vouloumanos, Onishi & Pogue, 2012; Vouloumanos Martin & Onishi, 2014). Infants are tuned to social cues, and could in principle already use these cues from the outset of word learning.

Because these two classes of theories—perceptual and social—are typically posed as mutually exclusive competitors, and because both are supported by compelling empirical findings, each has attempted to re-conceptualize evidence in favor of the other in its own terms. For example, researchers in the perceptual tradition have shown that 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 *because* infants infer that they are being presented by a social, pedagogically motivated caregiver (Csibra, 2010; Deligianni, Senju, Gergely & Csibra, 2011).

Unifying these two accounts within a single framework provides a promising theoretical alternative. As advocates of social word learning point out, 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, Bates, Casadio, Fenson, Fenson *et al.*, 1995; Tardif, Fletcher, Liang, Zhang & Kaciroti, 2008), young children also learn verbs, adjectives, and many other word types (Fenson, Dale, Reznick, Bates, Thal *et al.*, 1994; Clark, 2003; Bergelson & Swingley, 2013). It is likely that the kinds of mechanisms advanced in perceptual accounts of early word learning are not completely sufficient to explain this diversity of word meanings (Gleitman, 1990; Bloom, 2000; Waxman & Gelman, 2009), despite work showing that these mechanisms do play a role (Piantadosi,

Tenenbaum & Goodman, 2012; Scott & Fischer, 2012). While we focus here on concrete noun learning, the scalability of learning mechanisms to the broader vocabulary is an important theoretical concern for unifying models.

One possible proposal for unification is that 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 social cues are powerful predictors of a speaker's referential intentions. Similarly, children might reduce the weight they assign to perceptual cues as they discover that they are ineffective predictors of referential intention (Hollich, Hirsh-Pasek, Golinkoff, Brand, Brown *et al.*, 2000; Golinkoff & Hirsh-Pasek, 2006). The goal of the current work is to test this kind of unifying 'developmental cue combination' account; we begin by reviewing previous research supporting such an account and then derive several predictions, which we test in two experiments.

Data supporting developmental cue combination

The primary support for a developmental cue combination account comes from studies that pit perceptual salience against social information (e.g. speaker gaze) and measure children's learning under these conditions at different ages. Hollich *et al.* (2000) presented 12 studies that varied the referential cues highlighting two different objects in 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 these studies, 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. In a follow-up experiment, Pruden, Hirsh-Pasek, Golinkoff and Hennon (2006) found that 10-month olds did not attend to the speaker's

gaze at all when it was in competition with perceptual salience.

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

Together, these studies present data that are supportive of a developmental cue combination account, but nevertheless leave several questions unresolved. First, these studies present an incomplete picture of children's behavior in learning trials. While they describe children's preferential looking at one object over the other, they miss a third critical component of these naming events: attention to the speaker. While children's use of perceptual cues can be inferred straightforwardly from looks to the objects, characterizing their uptake of social cues requires measuring their engagement with the speaker. Second, it is unclear from prior work whether changes in the relative strengths of social and perceptual cues are due to increasing weights for social cues, decreasing weights for perceptual cues, or both. Answering this question would require comparing conditions where cues are in opposition to a condition where *only* the social cue is available.² The goal of our current studies is thus to gather a large, developmentally broad eye-tracking dataset on children's behavior in cue combination tasks, both during learning and at test. This dataset allows us to directly test predictions of the developmental cue combination account, in turn addressing these questions.

Testing predictions of developmental cue combination: the current study

Weighted cue combination is an intuitive, computationally simple model of the process of change in early word learning. Indeed, a number of computational models have implemented a version of this idea (Frank, Goodman & Tenenbaum, 2007; Yu & Ballard, 2007; Frank, Tenenbaum & Fernald, 2013). Developmental cue com-

² While Moore *et al.* (1999) did compare these conditions, they have relatively low power to detect differences between them due to small samples and a challenging test measure (forced choice responding). Hollich *et al.* (2000) ran studies using gaze alone with 12-month-olds, but not with older age groups.

bination 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). Despite this intuitive plausibility, a number of detailed predictions of cue combination models remain untested, specifically as they apply to these models as a description of developmental change in children's concrete noun learning.

In the current study, we test four predictions of the cue combination model of developmental change:

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

Prediction 1 is derived directly from previous work on the emergentist coalition model, which has suggested that a major developmental change during the second year was a move from reliance on salience to reliance on social cues (Hollich *et al.*, 2000; Golinkoff & Hirsh-Pasek, 2006). The developmental cue combination account predicts that this shift should be a major driver of developmental change in word learning during this period. Predictions 2 and 3 are corollaries of Prediction 1, again following from previous work that suggested that perceptual cues were less relevant for older children, whose learning was informed more by social cues. Finally, Prediction 4 comes from the idea that cue combination affects learning by biasing attention; thus, increasing weights on social cues relative to salience should be reflected in a developmental shift in the distribution of children's eye gaze when cues are in conflict.

In two experiments, we show that none of these predictions are 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. We end by discussing possible modifications to this view.

Experiment 1

In nearly all previous experiments investigating cue combination in early word learning, social cues were pitted against perceptual cues (see Moore *et al.*, 1999). Thus, 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, or (3) perceptual and social cues both change in their *relative* weight. Experiment 1 was designed to distinguish between these three possibilities by independently measuring the development of children's abilities both to follow and to 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.

We tracked children's eye movements 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 the child, 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, they were also tested in preferential looking trials with two known items. In Experiment 1, the two novel toys were chosen to be equally salient to children (see Appendix).

Method

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: parent-reported developmental issues (e.g. autism, language delays, etc.; $N = 27$), failure to calibrate ($N = 58$), and less than 75% exposure to English ($N = 36$).³ 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 3–3.5-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

³ These exclusion criteria were preset in this study on the basis of previous work (Yurovsky, Wade & Frank, 2013). Our high exclusion rate is due to inclusive recruitment, congruent with the outreach component of recruiting in a museum context.

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 child, and then turned towards one of the two toys on the screen, labeling it three times in a short monologue (Figure 1a). 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 to find the target of a novel label (*dax*). These ME trials were designed as a strong test of mapping formation. Looking to the correct target on Novel trials alone could result from familiarity or preference rather than mapping, but correct performance on both Novel and ME trials could only result from knowledge of the specific label used during the learning phase.

Procedure

We collected eye-movement data with an SMI RED corneal-reflection eye-tracker mounted on an LCD monitor, sampling at 120 Hz. 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 the test trials, in which their knowledge of both familiar and novel word-object mappings was tested. The entire experiment consisted of four learning trials, eight *Familiar*, six *Novel*, and six *ME* test trials. We additionally inserted 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 (see below).

Results and discussion

We analyzed children's eye movements using a Region 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 learning and test behaviors were quantified by measuring their proportion of looking to each ROI on each trial. To ensure that proportions were



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 ~12s. 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.

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. To ensure accuracy in our identification of point of gaze, 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 correct the calibration of eye movements for all learning and test trials (Frank, Vul & Saxe, 2012).

We begin by describing children's behavior during both learning and test trials, then we present statistical analyses, and finally discuss implications for the predictions of the developmental cue combination model.

Descriptive analyses

First, children in all age groups spent the majority of learning trials fixating the speaker's face, although this

proportion decreased across development. Interestingly, looking to the face remained relatively constant over the course of the naming events (Figure 2). Similarly, the second most-fixated ROI at all ages was the target object. Proportion of looks to the Target increased across early development, and like Face-looks, remained relatively constant over the naming event. Looking to the competitor was consistently low at all ages. Thus, across development, children appeared to attend to the speaker, and follow her gaze to the target object over the competing object. The major developmental change during the learning phase appeared to be improvement in the ability to disengage from the speaker.

At test, children at all ages were able to find the target referents of known words on Familiar trials. Both speed and accuracy improved across development, and additionally the 3- and 3.5-year-olds attended to the target for longer, rebounding after checking back to the competitor (Figure 3). Looking patterns were broadly similar on Novel trials, although overall accuracy was lower and the youngest age groups appeared to find the

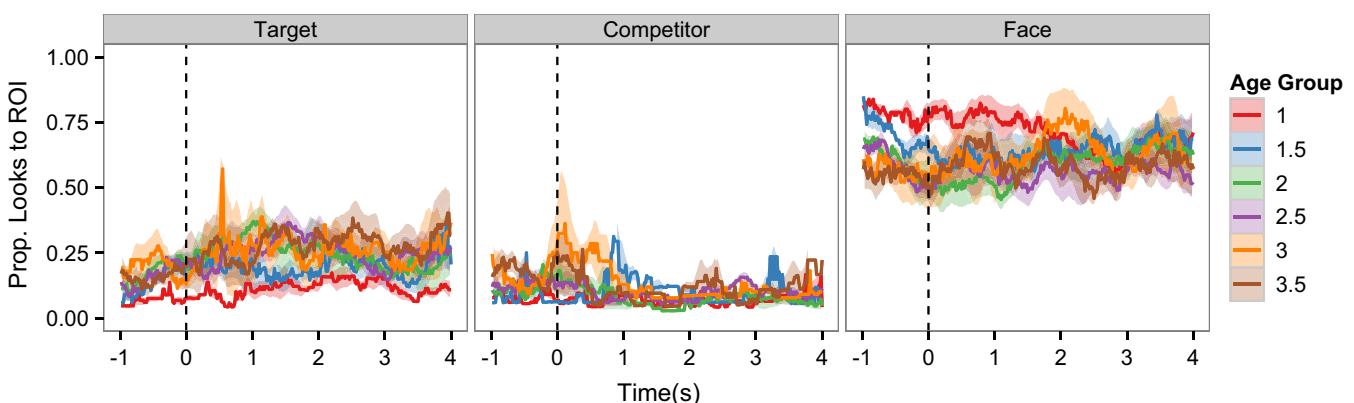


Figure 2. Looking behavior during learning trials in Experiment 1. Children in all age groups spent the majority of naming events looking at the speaker's face, but looking to the face decreased and looking to the target increased across development. Looking to the competitor stayed constant. Each curve represents an age-group mean, and shaded regions indicate ± 1 SE. For clarity, curves were smoothed by taking the average within a 60 ms moving window.

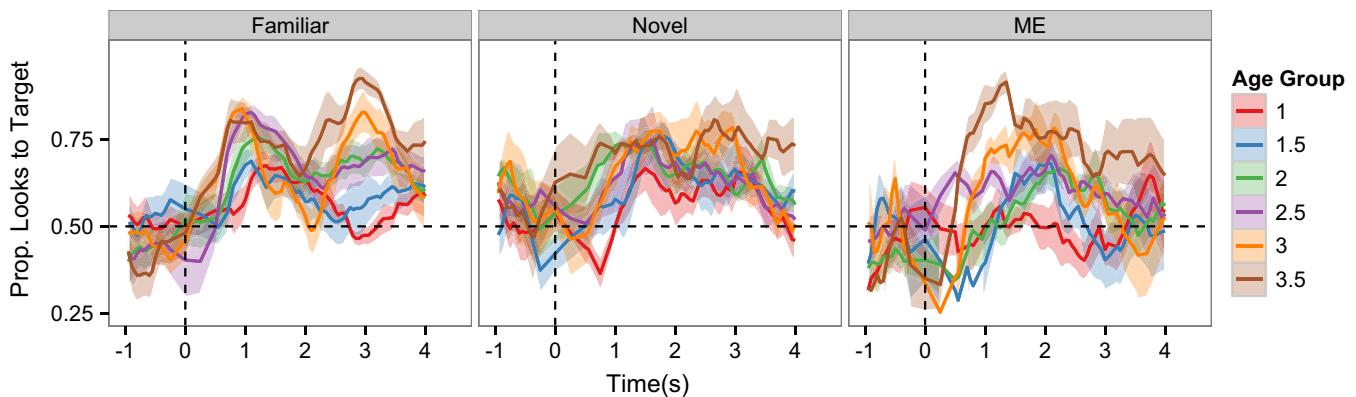


Figure 3. Looking behavior at test in Experiment 1. Children improved in both speed and accuracy on all three trial types—Familiar, Novel, ME—across development. Plotting conventions are consistent with Figure 2.

target more slowly. Mutual Exclusivity trial performance also showed similar trends, although there was little evidence that the youngest children—1–1.5-year-olds—were correctly looking at the previously un-named object on ME trials (consistent with prior work on mutual exclusivity with this age group; Halberda, 2006).

A clearer way of seeing these trends is through onset-contingent analyses (Fernald, Zangl, Portillo & Marchman, 2008). Figure 4 shows the same data split by the object that was the focus of attention at the onset of the label. In each panel, the two lines show the proportion of children over time who switched to look at other object if they—by chance—began the trial on the target (solid), or competitor (dotted). If children knew the correct referent on a trial and looked at it, those who began on the competitor should have switched to the target, and those who began on the target should have remained on the target. The area between these two curves is informative about the strength of children’s discrimination of target vs. distractor, and the time point at which curves diverge is informative about the speed with which the discrimination was made. In addition, these trajectories can be informative about the processes underlying this discrimination (Halberda, 2006).

These graphs show broad improvement both in speed and accuracy across development; the area between the two lines in each curve is larger, and the point of divergence between the two curves is earlier. They also showcase the difference in difficulty between Novel and Mutual Exclusivity trials. The two ME curves for 1-year-olds overlap for essentially the entire duration of test trials; for 1.5-year-olds, the dotted line is above the solid line (indicating fixation to the correct object) for only a short window of time in the middle of the trial.

Together, these descriptive analyses suggest that from the earliest age we measured, children were engaged in

the naming events and followed the speaker to the target of her reference using the direction of her head and gaze. Across development, children became better able to disengage from the speaker and spent more time looking at the target. In addition, speed and accuracy on all three test types (Familiar, Novel, and ME) improved across development. To quantify these impressions with standard measures, we aggregated over these time course measures to compute proportion of looking in a particular temporal window of interest. These windows began at the point of disambiguation for each trial. For test trials, the point of disambiguation was the onset of the target label, and for learning trials it was the rotation of the speaker’s head. 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.

Statistical analyses

Inspection of time course plots showed that children in all age groups were successful at attending to and following the speaker’s social gaze during learning trials. Statistical analysis confirmed this impression: children of all ages spent more time looking at the target than at the competitor during learning trials (smallest $t(23) = 3.20$, $p < .01$, $d = .65$) (Figure 6). 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 5).⁴

⁴ All data and code for analysis available at <http://github.com/dyurovsky/ATT-WORD>.

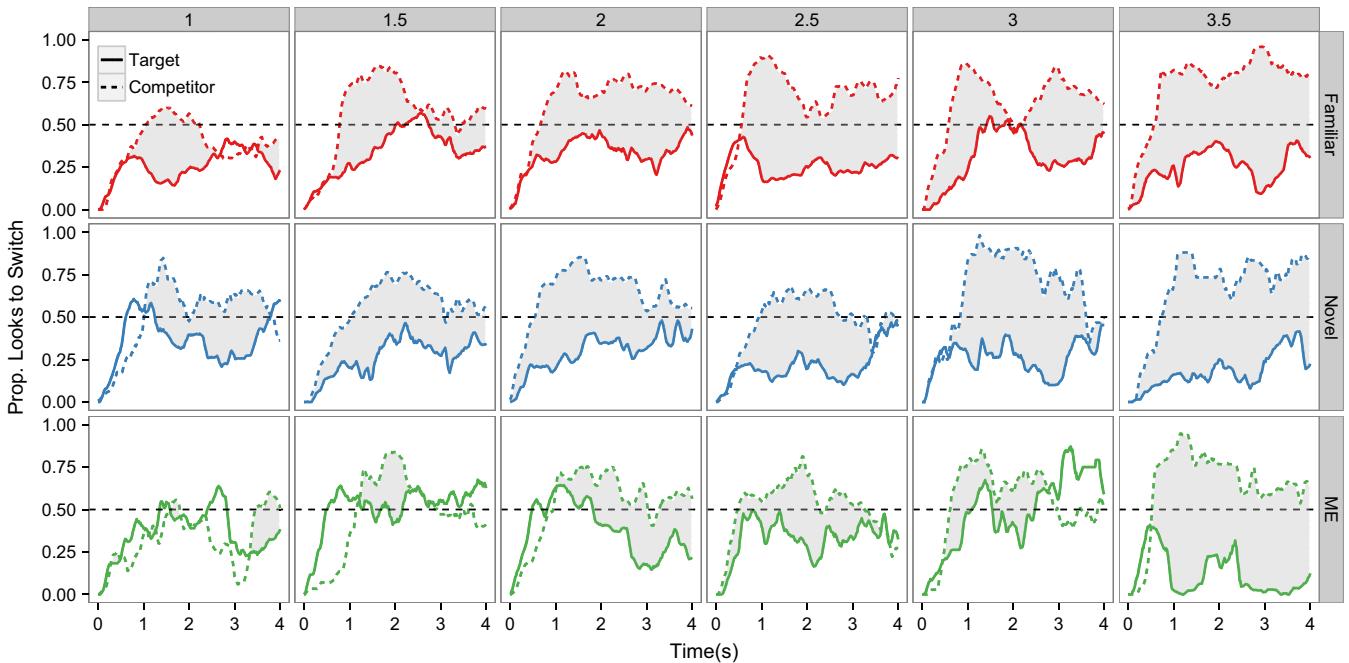


Figure 4. Onset-contingent plots for test trials in Experiment 1. Each panel shows the proportion of children who shifted from one referent to the other across time. The solid line represents children who (by chance) began the trial on the correct target, and the dotted line represents children who began on the competitor. The size of the gray shaded region—the region where more switches were made from competitor-to-target than from target-to-competitor—is indicative of how much children discriminated between the two toys in the context of the label.

Similar 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 ($t(26) = 1.65, p = .11, d = .32$), 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 $t(17) = 2.10, p = .05, d = 1.16$). These analyses are consistent with previous work showing the emergence of success on ME trials from approximately 16–18 months (e.g. Halberda, 2006; Bion, Borovsky & Fernald, 2013).

To quantify the change we observed across development, we fit a mixed effects logistic regression to the proportion of looking data during both learning and test (Jaeger, 2008). Our regression model predicted the proportion of looking to the target on each trial using trial type (Learning, Familiar, Novel, ME) with Novel as the reference trial type. To make data from the learning phase comparable with those from the test phase, we computed the proportion looking to the target vs. the competitor object, excluding looks to the face. This analysis revealed significant improvements in target looking across age ($\beta = .61, z = 4.03, p < .001$), as well as a significant effect of Learning as compared to Novel trials ($\beta = 1.18, z = 3.11, p < .01$). No other trial types

were significantly different from the Novel baseline, and a model with an interaction between age and trial type was not a significantly better fit to the data. These statistical analyses together suggest that children were more successful at finding the correct target during learning trials, but that they improved similarly on all trial types across development, capturing the general developmental trends shown in Figure 6.

Summary

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 in this skill across development. Further, improvements in gaze-following paralleled improvements in both finding the referents of novel words on subsequent test trials, and also finding the referents of familiar words.

How well do these results fit with the developmental cue combination account? First, there was no conflict between cues in this experiment, so any changes observed in behavior could not be due to *relative* re-weighting of cues (Prediction 1). Thus, this prediction is not supported. One way around this conclusion would be to argue that what we observed was developmental

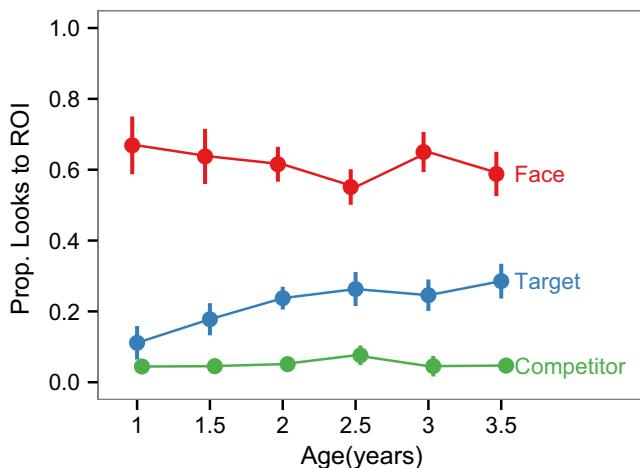


Figure 5. Proportion of children's looking the target toy, competitor toy, and the speaker's face during learning trials in Experiment 1, summarizing the developmental trends shown in Figure 2. Children of all ages spent the majority of the learning trials looking at the speaker's face. Disengaging from the face and fixating the target increased across development. Error bars indicate a 95% confidence interval computed by non-parametric bootstrap.

change in *absolute* weight for social information. But such a description seems to be at best a misleading account of the data: Performance on familiar word recognition trials appeared to be changing at roughly the same rate as recognition of novel words. This congruence

suggests that a more parsimonious account of the observed development would be general changes in speed of word recognition (e.g. Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998; Fernald, Perfors & Marchman, 2006), not changes in social cue weight, which after all would be irrelevant for familiar word recognition.

Second, these results also speak against Prediction 4 of the cue combination account, that cue weights drive attention during learning. Children at 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 we observed was the ability to *disengage* from the speaker's face. This result is congruent with findings from 1–2-year-olds suggesting developmental increases in the ability to disengage from faces in favor of other targets (Frank *et al.*, 2012). Disengagement in our stimuli was due to gaze-following—and hence relevant to the use of social information—but in some sense this behavior is the opposite of what the cue combination account predicts. The less children's attention is captured by particular social stimuli, the more they are free to attend flexibly to aspects of the context that may be relevant for understanding.

In sum, data from Experiment 1 did not support the predictions of the developmental cue combination account. Instead, the developmental trends we observed appeared more consistent with a different account. In

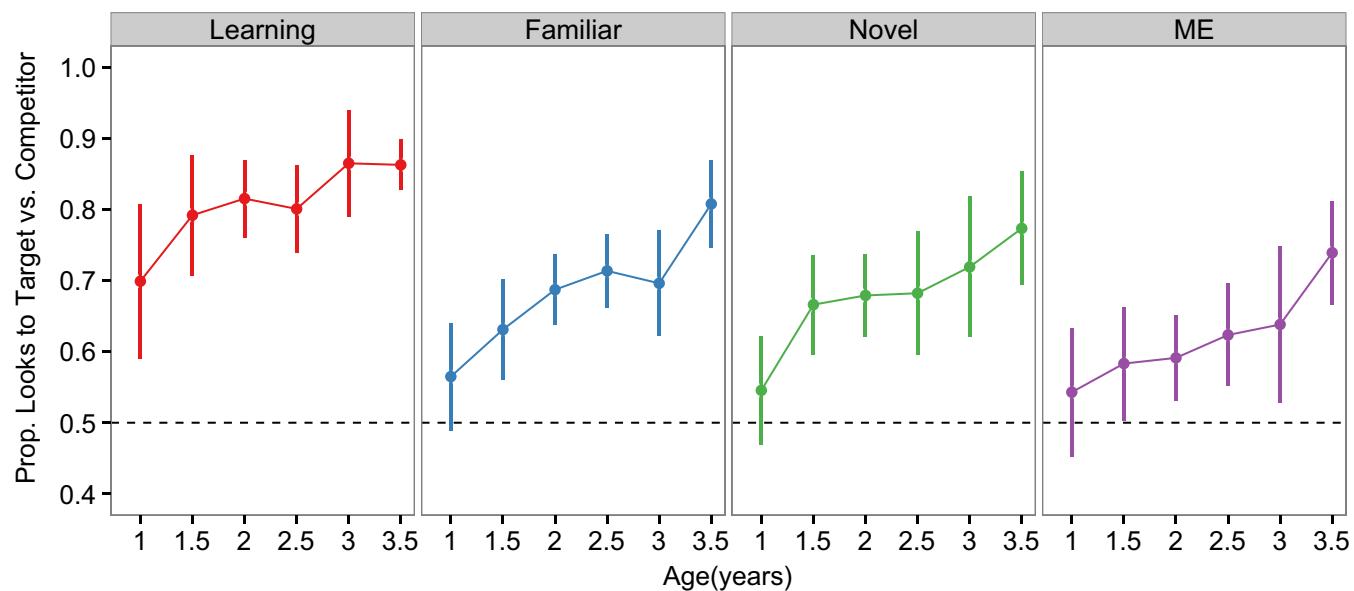


Figure 6. Proportion of time children fixated the correct target on each trial type in Experiment 1. Children improved on all measures across development. Each dot indicates one half-year age group and each line represents a 95% confidence interval computed by non-parametric bootstrap. A proportion of .5 indicates chance performance.

our data, developmental changes appeared to be driven by changes in general processes of attention and memory combined with increasing familiarity with and fluency in processing spoken language. Nevertheless, the primary data of interest for earlier tests of cue combination were conflicts between cues (Hollich *et al.*, 2000), so to fully test these theories it is important to investigate such conflicts. In Experiment 2 we manipulated the relative salience of the target and competitor objects during learning trials with gaze cues. This manipulation allowed us to measure how salience affects children's looking during both learning and test.

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 (see Appendix). For children in the *Salient* condition, the target was the more interesting toy, and the competitor the less interesting toy. In the *NonSalient* condition, the identities of the toys were switched—the target was the less salient toy. Experiment 2 thus allowed us to investigate children's use of social cues to learn new words when 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: 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 *NonSalient* condition, data were collected from 126 children, 71 of whom were excluded for one or more of the following reasons: 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

To determine the effect of perceptual salience on word learning, we compared children's looking in the *Salient* and *NonSalient* conditions not only to each other, but also to the *Balanced* condition tested in Experiment 1. As before, we begin by presenting basic descriptive analyses, then statistical analyses, and finally a summary of implications for the naïve cue combination account of early word learning.

Descriptive analyses

As in Experiment 1, children spent the majority of learning trials looking at the speaker's face, and this proportion remained relatively constant over the course of the naming event. Further, children again successfully followed the speaker to the correct target—both when it was the *Salient* toy and when it was the *NonSalient* toy—in all three age groups. Figure 7 shows the data from the *Salient* and *NonSalient* conditions alongside data from the *Balanced* condition of Experiment 1. Looking behavior across all three conditions (columns) appears strikingly similar, showing very little effect of perceptual salience during learning trials. We return to this observation in the statistical analyses below.

As expected, looking behavior during Familiar test trials was similar across the experiments. Behavior on Novel and ME trials showed a strong effect of condition, however (Figure 8). On Novel trials, children in all age groups performed best in the *Salient* condition, worst in the *NonSalient* condition, and at an intermediate level in the *Balanced* condition. That is, children were faster and more accurate at finding the target toy at test when it was the perceptually salient object. In contrast, the Mutual Exclusivity trials showed the opposite pattern: Children in all age groups appeared to succeed in finding the previously unnamed salient toy when they heard an unfamiliar label, but they had difficulty in the same task when the unnamed toy was not salient. That is, salience appears to have had essentially the same effect on looking behavior regardless of trial type or age: Children were more likely to look at the salient toy.

The effect of salience on looking behavior is even clearer in the onset-contingent analysis (Figure 9). Across age groups, area between the target and

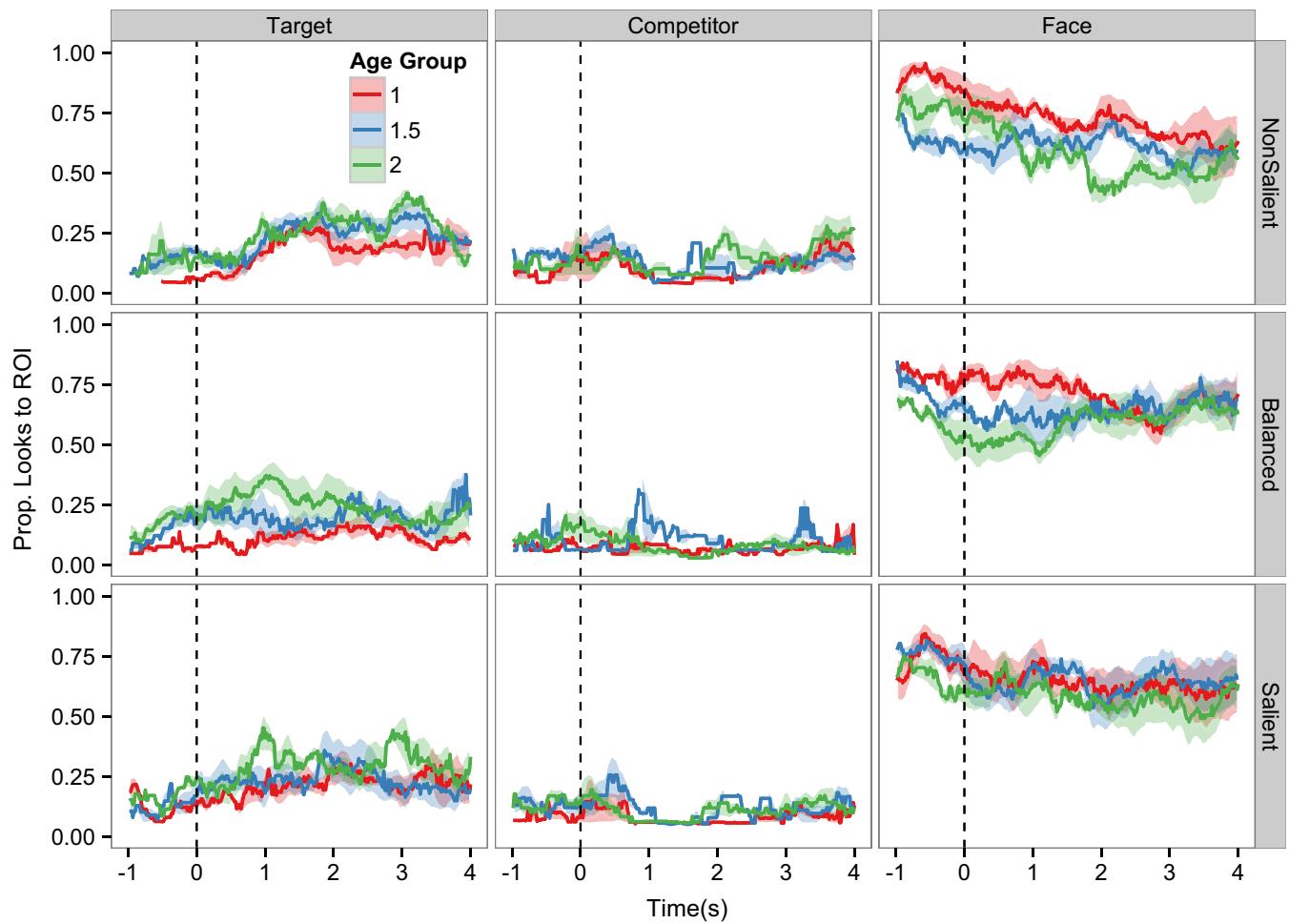


Figure 7. Looking behavior during learning trials in both the Balanced condition of Experiment 1 and the Salient and NonSalient conditions of Experiment 2. In all conditions, children in all age groups were successful in following the speaker's gaze to the target. However, all children also spent the majority of learning trials fixating the speaker's face. Plotting conventions are as in Figure 2.

competitor curves increases for Novel trials in the Salient condition and decreases for the NonSalient condition. Similarly, area increases on ME trials for the NonSalient condition and decreases for the Salient condition.

Statistical analyses

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 10, top). As in Experiment 1, children of all ages spent more time looking at the target than the competitor (smallest $t(14) = 3.53, p < .01, d = 1.32$). Looking time to both target and distractor again made up the minority of their dwell time, however; children spent the majority of learning trials looking at the speaker's face (smallest

proportion: 2-year-olds in the NonSalient Condition: .51).

In principle, this result could be due to the toys being too similar in their salience, making our experiment a weak test of the cue combination model. But in fact, the differing salience of the two objects exerted a very strong effect during 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 $t(19) = 2.96, p < .01, d = .66$). When the target was nonsalient, no age group showed evidence of learning on Novel test trials (largest $t(13) = 1.46, p = .17, d = .39$). 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

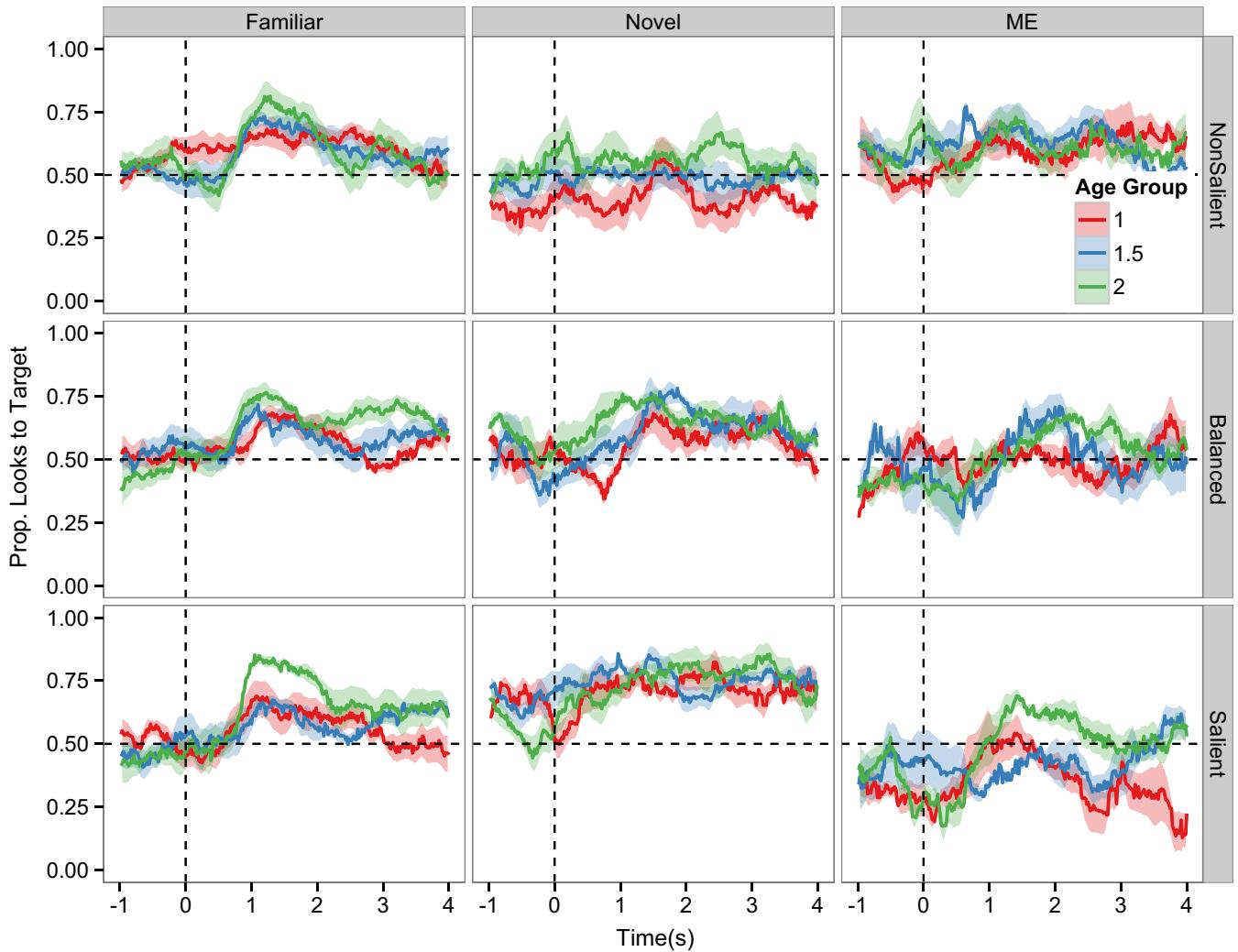


Figure 8. Looking behavior at test in all conditions. In contrast to learning trials, salience had a large effect on looking at test. At all ages, children were drawn to the salient toy regardless of trial type. Plotting conventions are consistent with Figure 2.

competitor) at below chance levels (smallest $t(20) = -2.29$, $p < .05$, $d = .50$). In the NonSalient condition, even the youngest children looked at the correct referent on ME trials at above chance levels (smallest $t(22) = 4.51$, $p < .001$, $d = .94$). Figure 10 shows a summary of looking behavior across both Experiments 1 and 2.

The effect of perceptual cues at test did not appear to change across the age range we tested in Experiment 2. We fit a mixed-effects logistic regression to the data from both experiments to determine how age and experimental condition affected 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. No models with additional interactions improved fit to the data, indicating that neither condition nor trial type interacted with age (Table 1).

Summary

These analyses again suggest a model of early word learning different in a number of ways from naïve cue combination. First, the perceptual cue had a consistent effect across development, attracting attention similarly for children in all age groups (contra Prediction 3). Second, this effect was largely absent during learning trials, suggesting that perceptual salience of the target and competitor did not affect children's attention during the naming events (contra Prediction 4).

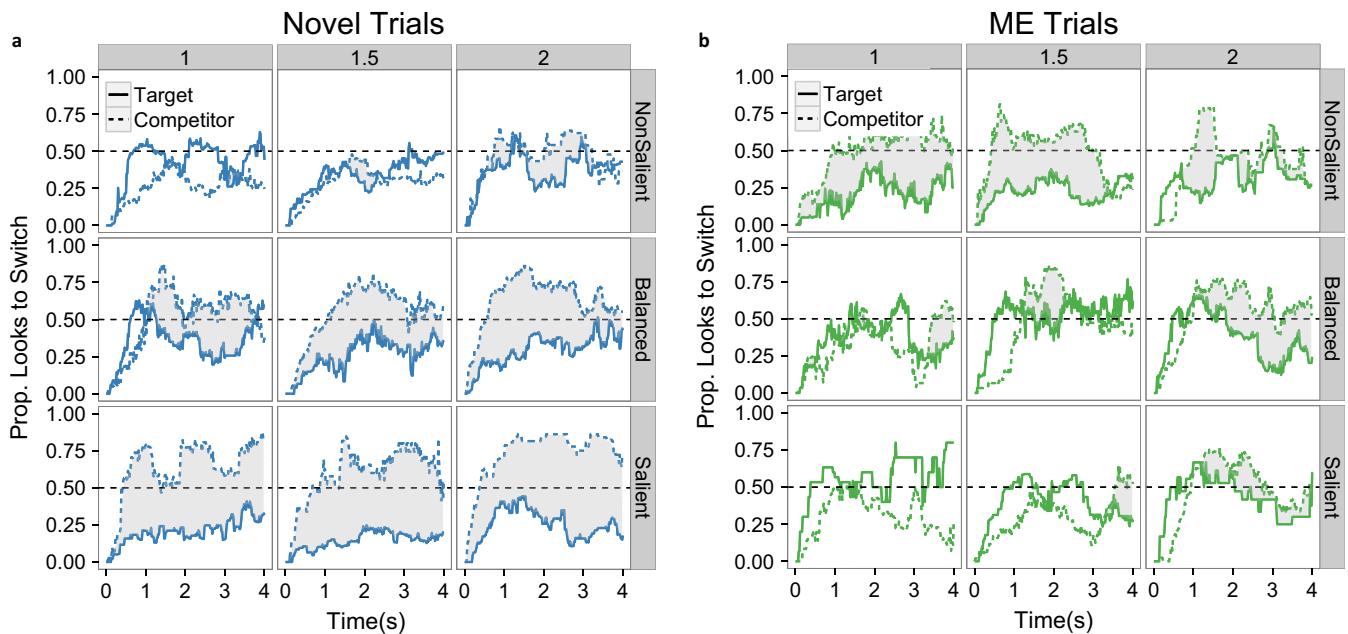


Figure 9. Onset-contingent plots for test trials in all conditions. The effect of salience is apparent in comparing the size of the discrimination area for the same trial type across conditions. Areas increased for Novel trials in the Salient condition and decreased in the NonSalient condition. The opposite pattern appears for ME trials. Plotting conventions are consistent with Figure 4.

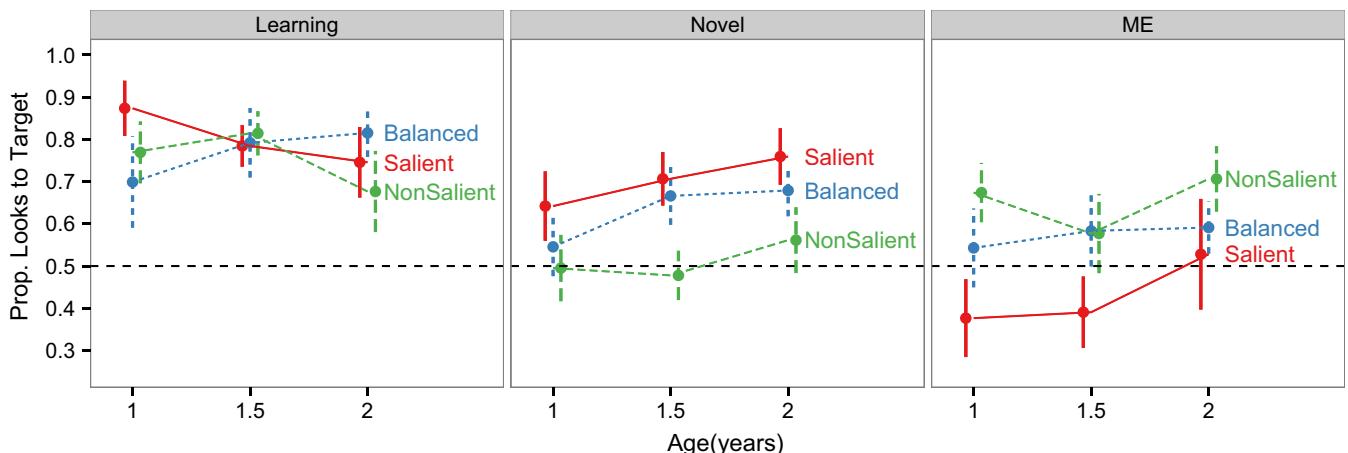


Figure 10. Proportion of time children fixated the correct target on Learning and Test trials in Experiments 1 (three youngest age groups) and 2. Salience had the predicted effect on looking behavior at test, but relatively little during learning. Each dot indicates one half-year age group and each error bar represents a 95% confidence interval computed by non-parametric bootstrap. A proportion of .5 indicates chance performance.

As in Experiment 1, older children in both the Salient and NonSalient conditions were better at finding the target on both learning and test trials. But critically, this difference was unrelated to the effect of perceptual salience. Further, after controlling for performance on Familiar test trials, even children's age failed to reach significance as a predictor of looking to the target on

learning and test trials. These findings strongly suggest that, although there is clear change in word learning across development, this development is not driven by changing cue weights (Prediction 1). Instead, the difference appears to derive from changes in other cognitive processes—attention, memory, and language processing.

Table 1. Mixed-effects logistic regression model for looking behavior in Experiments 1 and 2. The model was specified as Prop.Correct ~ Age + Condition * TrialType + Familiar + (1 | Subject) with the Novel test type in the Balanced Condition as the reference category

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
NonSalient × Learning	1.15	0.65	1.76	0.08 .
Salient × ME	-2.23	0.61	-3.65	<.001 ***
NonSalient × ME	1.59	0.54	2.92	<.001 **

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 and parsimonious framework to unify social and perceptual factors in early word learning (Hollich *et al.*, 2000; Yu & Ballard, 2007; 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). Perceptual salience exerted its effects mostly at test, and did so consistently across early development instead of decreasing in weight (contra Prediction 2). Social cues appeared to have a differential effect across development, but changes in learning from social cues mirrored increases in familiar word recognition, suggesting that the underlying cause might not be change in cue weight but changes in more general cognitive factors (contra Prediction 3). Finally, developmental changes during learning appeared to be driven by disengagement from the social stimulus, not disengagement from the perceptually salient competitor object (contra Prediction 4).

The naïve cue combination model provides important insights about the different constructs involved in early word learning, but our data nevertheless cause us to reject it as a broader framework for two reasons. First, in consideration of other evidence on early social word

learning, we find cue combination to be too impoverished a framework to accommodate the range of communicative inferences that have been shown in young children. Second, naïve cue combination fails to provide an adequate account of the developmental changes we observed in our experimental data.

From cue weighting to communicative inference

Stepping back from the stripped-down mapping paradigm we studied here, there is a broad literature attesting to the remarkable inferences that early word learners can make on the basis of social evidence. For example, Tomasello (2000) lists an impressive variety of different inferences from the same limited materials—a speaker's reference in context, modulated by low-level cues such as gaze and pointing as well as high-level cues such as discourse context or even a speaker's admission of an error. While these inferences are primarily attested with children 18 months and older (e.g. Baldwin, 1993; Akhtar, Carpenter & Tomasello, 1996), both observational analyses and looking-time paradigms provide evidence that the same information is present and used in interpretation for even younger children (Frank *et al.*, 2013; Vouloumanos *et al.*, 2012, 2014). These findings simply do not admit to an interpretation in terms of a linear combination of cues, whatever their weights. Instead, cues are interpreted depending on their context and the learner's estimation of the speaker's intentions.

In addition to the issues of applying cue combination to single learning instances, the cue combination framework does not speak to the challenges of interpreting across situations. Word learning often 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 across 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 presented 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 interpretation in the moment. While we do not develop such an account here, we believe that this is a critical next step for theory building.

Domain-general change as an account of developmental differences in word learning

While other data speak against cue combination as a broader model of social word learning, one attraction of the cue combination model is the clarity of its predictions about the changes that should be observed in word learning across early development, as well as the learning input that should produce these changes. This predictive clarity was the primary rationale for our focus on cue combination in the present studies. Unfortunately, these predictions were not borne out in our data.

The factors involved in predicting children's responses in our dataset were indeed those identified by cue combination models: both social information and salience played a strong role in determining behavior. But changes in the weights on these cues did not appear to account for our data. At test, perceptual salience was a strong driver of even the oldest children's attention; and during learning, social information directed even the youngest children's attention. What changed, however, was whether the children *disengaged* from the social cue, not whether they engaged with it, and whether they *remembered* the mapping that they had learned, not whether they attended to it when it was more salient.

The nature of children's improvement in identifying the speaker's referent over development cannot be attributed to their learning the predictive validity of gaze as a referential cue. Instead, developmental changes in domain-general cognitive processes appear responsible, although our experiments were not designed to disentangle exactly what these were. Early childhood is a time of substantial development in attention and inhibition (Smith, Thelen, Titze & McLin, 1999; Diamond & Doar, 1989), memory (Cowan, 1997), and speed of processing (Kail, 1991; Dougherty & Haith, 1997). All of these cognitive processes play a role in our task. Flexible shifting of attention is critical for the learning phase, as is having the inhibitory control to disengage with the face; at test, similarly, the ability to disengage from a more salient competitor image is critical. Memory for the mapping between word and object should play a large part in performance at test as well (Horst & Samuelson, 2008). Finally, changes in speed of processing—whether for language specifically or overall—may underly some changes in performance at test, as has been suggested in familiar word processing (Fernald *et al.*, 1998; Fernald & Hurtado, 2006). Thus, our findings are congruent with a theoretical model proposed by Yu and Smith (2012b): Whatever the precise learning mechanisms underlying word-object mapping, these mechanisms must be supported by general processes of attention and memory.

Disentangling the roles of these distinct constructs will thus be an important goal for future work.

Further implications and conclusions

Despite these differences in interpretation from the cue combination account, our data are not generally inconsistent with the empirical results that motivated previous theory. As observed in previous studies, looking at the target object relative to the competitor object during learning trials increased steadily across development (Moore *et al.*, 1999; Hollich *et al.*, 2000). Similarly, novel word learning improved at a similar rate across development. In contrast to previous work, however, we were able to measure children's attention not just to the target and competitor toys, but also to the speaker. Using continuous eye-tracking measures thus gave us additional insight into the dynamics of children's attention and information processing, and ultimately allowed us to make different inferences from the data we observed.

In addition to its theoretical consequences, our work here has significant implications for users of two-alternative preferential looking displays. Both Reznick (1990) and Fernald *et al.* (2008) highlight the importance of matching targets in two-alternative displays. Our work provides further justification for these recommendations. When alternatives were not matched for perceptual salience, the relatively more salient object dominated children's looking preferences for all age groups at test. 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 less salient and exaggerated if the target was more salient. This overshadowing-by-salience was, if anything, more pronounced for the mutual exclusivity trials. Especially for 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.

In sum, early word learning is a fundamentally social process. The virtue of cue combination accounts is that they provide an important decomposition of the broader construct of social context into smaller components that allows for prediction and testing. However, the developmental predictions of such accounts are not supported in our data. Nevertheless, we believe that the kind of tentative synthesis that we have offered here—that social inference must be united with domain-general developmental change—is broadly consistent with the spirit of the Emergentist Coalition Model. It attempts to explain

both children's early social inferences and the often fragile nature of these inferences as well as their sensitivity to, and sometimes complete moderation by, lower-level perceptual processes. Children's social information processing is built from and operates on top of mechanisms for directing their attention, encoding auditory and visual information, retrieving this information later, and combining it across multiple contexts. We propose that the kind of developmental change in social information processing we observed in our data, and in the data used to motivate cue combination accounts, emerges not from changes in children's cue weights, but instead from developmental changes in these lower-level processes.

Acknowledgements

We are grateful to Janelle Klaas for collecting the data, to Carissa Romero for help in stimulus creation, and to Molly Lewis and all of the members of the Language and Cognition Lab for their feedback on this project. We also thank Heather Bortfeld and three anonymous reviewers for their careful and constructive comments on this manuscript. 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.

Appendix

In order to measure the impact of object salience on children's looking and learning, we needed four distinct objects, two that were equally salient (Balanced 1, Balanced 2), and two that were differentially salient (Salient, NonSalient). We began by estimating the relative salience of a number of different toys using aggregate adult judgments. Thirty-eight adults on Amazon Mechanical Turk were shown two toys 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 ~ 7.6 responses for each pair of toys. From these responses, we chose the two toys that were best balanced against each other as the Balanced toys. Two unbalanced toys were chosen as the Salient and NonSalient toys.

We then validated these relative saliences with a separate sample of children recruited from San Jose Children's Discovery museum. Demographic and exper-

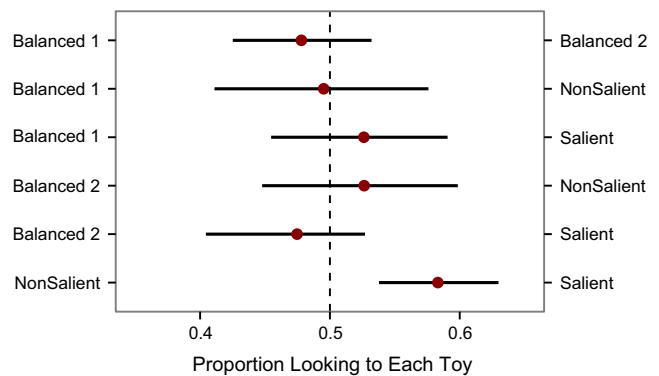


Figure A1. Children's preferential looking for all pairwise comparisons between objects in the Norming experiment. Points shows mean proportion of looking, and error bars show 95% confidence intervals computed by non-parametric bootstrapping. Children showed no pairwise preferences except for between the Salient and Non Salient toys.

imental data were collected from 33 children, 2 of whom were excluded for failure to calibrate. The final sample consisted of 8 1–2-year-olds (3 girls), 8 2–3-year-olds (3 girls), 9 3–4-year-olds (5 girls), and 6 4–5-year-olds (4 girls).

Children were shown each of the four toys used in Experiments 1 and 2, two at a time. Each trial was visually identical to the test trials from the main experiments, but instead of hearing a toy's label, children were only directed by the speaker to 'look!' Each toy was tested against each other toy twice, once in each possible left-right position. Children's proportion of looking to each toy on each trial was computed, and both trials for each pair were averaged together. Individual trials were dropped from analysis if children did not look at the screen for at least 50% of their duration.

Proportion of looking for each pair of toys was compared to chance with a one sample *t*-test. Children showed no significant preferences between any of the toys (largest $t(22) = .80$, $p = .43$, $d = .17$) except for between the Salient and NonSalient toys ($t(22) = 3.66$, $p < .01$, $d = .65$). Figure A1 shows mean looking proportions and 95% confidence intervals for each pairwise comparison. Thus, we can conclude that the Salient toy was indeed more salient than the NonSalient toy to children, and that the two Balanced toys were indeed balanced for object salience.

References

- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, **67**, 635–645.

- Baldwin, D.A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, **20**, 395–418.
- 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.
- Bion, R.A.H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, **126**, 39–53.
- 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., et al. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, **10** (2), 159–199.
- Clark, E.V. (2003). *First language acquisition*. New York: Cambridge University Press.
- Cowan, N. (Ed.) (1997). *The development of memory in childhood*. Hove: Psychology 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.
- Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology*, **22** (3), 271–294.
- Dougherty, T.M., & Haith, M.M. (1997). Infant expectations and reaction time as predictors of childhood speed of processing and IQ. *Developmental Psychology*, **33** (1), 146.
- 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. et al (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59**, (5, Serial No. 242).
- Fernald, A., & Hurtado, N. (2006). Names in frames: infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, **9**, F33–F40.
- Fernald, A., Perfors, A., & Marchman, V.A. (2006). Picking up speed in understanding: speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, **42**, 98–116.
- 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.
- Fernald, A., Zangl, R., Portillo, A.L., & Marchman, V.A. (2008). Looking while listening: using eye movements to monitor spoken language. In I.A. Sekerina, E.M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: Online methods in children's language processing* (pp. 113–132). Amsterdam and Philadelphia, PA: John Benjamins.
- 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 J.C. Platt, D. Koller, Y. Singer, & S. Roweis(Eds.), *Advances in neural information processing systems, volume 20* (pp. 1212–1222). Cambridge, MA: MIT Press.
- 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.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, **1**, 3–55.
- 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*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, **53**, 310–344.
- Hollich, G.J., Hirsh-Pasek, K., Golinkoff, R.M., Brand, R.J., Brown, E. et al. (2000). Breaking the language barrier: an

- emergentist coalition model for the origins of word learning. *Monographs of the Society of Research in Child Development*, **65** (3), i–vi, 1–123.
- Horst, J.S., & Samuelson, L.K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, **13**, 128–157.
- 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.
- Kail, R. (1991). Processing time declines exponentially during childhood and adolescence. *Developmental Psychology*, **27**, 259–266.
- 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*, **123**, 199–217.
- 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.
- Reznick, J.S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics*, **11** (02), 145–166.
- 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: Oxford University Press.
- Smith, L.B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: the task dynamics of the a-not-b error. *Psychological Review*, **106** (2), 235–260.
- Tardif, T., Fletcher, P., Liang, W., Zhang, Z., & Kaciroti, N. (2008). Baby's first 10 words. *Developmental Psychology*, **44**, 929–938.
- Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, **10**, 401–413.
- Vouloumanos, A., Martin, A., & Onishi, K.H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, **17**, 872–879.
- 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, USA* **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 Sciences*, **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., & Ballard, D.H. (2007). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing*, **70** (13), 2149–2165.
- 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.

Received: 13 August 2014

Accepted: 17 June 2015