Semantic Specificity in One-Year-Olds' Word Comprehension

Elika Bergelson & Richard Aslin

Published online: 30 Jun 2017.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Check for updates

Semantic Specificity in One-Year-Olds' Word Comprehension

Elika Bergelson [a,b] and Richard Aslin [a]

aBrain and Cognitive Sciences, University of Rochester, Center for Language Sciences; bPsychology & Neuroscience Department, Duke University

ABSTRACT
The present study investigated infants' knowledge about familiar nouns. Infants (n = 46, 12–20-month-olds) saw two-image displays of familiar objects, or one familiar and one novel object. Infants heard either a matching word (e.g. "foot" when seeing foot and juice), a related word (e.g. "sock" when seeing foot and juice) or a nonce word (e.g. "fep" when seeing a novel object and dog). Across the whole sample, infants reliably fixated the referent on matching and nonce trials. On the critical related trials we found increasingly less looking to the incorrect (but related) image with age. These results suggest that one-year-olds look at familiar objects both when they hear them labeled and when they hear related labels, to similar degrees, but over the second year increasingly rely on semantic fit. We suggest that infants' initial semantic representations are imprecise, and continue to sharpen over the second postnatal year.

Introduction

In the first two years of life, infants acquire their language-specific phonology, and begin to populate their receptive and productive lexicons. Indeed, months before their first birthday, infants show modest but consistent comprehension for common and proper nouns that are frequent in their daily input (Bergelson & Swingley, 2012, 2015; Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999, 2012). In the subsequent months, comprehension improves, and production begins (Fenson, Dale, Reznick, & Bates, 1994; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). However, the nature and bounds of infants' early lexical categories remains largely unclear, especially in the domain of meaning.

Part of understanding a word involves making appropriate generalizations, in both the sound and meaning domains. In the sound domain, for instance, infants must deduce that "tog" is not an acceptable way to say "dog"; this phonological learning process is part of developing appropriate word-form specificity. Previous research has found that by around the first birthday, infants know the precise sounds that make up common words, e.g., looking less at a cup when hearing "kep", than when hearing "cup" (Mani & Plunkett, 2010; Swingley & Aslin, 2002).

In the present study, we examine the analogous question of *semantic* specificity. Our central question is whether infants appropriately constrain what "counts" as a referent for words they know, in the semantic domain, during early word comprehension. More concretely, we compare two alternatives: (1) infants' lexical representations are overly inclusive (such that in the presence of a dog, hearing either "cat" or "dog" equivalently triggers dog-looking); and (2) infants' lexical representations are appropriately bounded (such that hearing "dog" elicits more dog-looking than hearing "cat" does).

Classic work on language production (Barrett, 1978; Rescorla, 1980) provides compelling evidence for lexical overgeneralization (e.g., labeling a lion as "cat"), but evidence for overextension in comprehension (especially before production begins) is far more limited. In toddlers, Naigles, Gelman, and colleagues have found that overextensions in production do not necessarily reflect semantic representations in comprehension (Gelman, Croft, Fu, Clausner, & Gottfried, 1998; Naigles & Gelman, 1995). In that work, while 2–4-year olds generated some examples of "true overextensions" in comprehension, larger age-related differences emerged for production. Here we investigate semantic representations in infants' lexicons over the second year of life, when word production is still quite limited.

Understanding early semantic specificity is critical for constructing learning theories for lexical acquisition. A learner that only accepts instances that are very closely matched to her experiences with a word and referent has a very different task before her in making appropriate generalizations to novel instances than one who starts off overly accepting, and then winnows down to appropriate extensions of a word. The present study looks for evidence of the latter in 12–20-month-old infants.

Early word comprehension & semantic knowledge

Previous studies of early word comprehension have generally measured infants' knowledge in one of several ways: by labeling one of several co-present visual referents while measuring where infants look on a screen (, 2013a, 2015; Fernald et al., 1998; Fernald, Zangl, Portillo, & Marchman, 2008; Tincoff & Jusczyk, 1999, 2012), by presenting infants with a single image that either matches or mismatches the name they hear while measuring looking time or EEG signals (Parise & Csibra, 2012), or by having parents judge what words they think their infant understands (Dale & Fenson, 1996). Such tasks are undertaken with either highly familiar words (Bergelson & Swingley, 2012; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999, 2012) or with novel words to assess mutual exclusivity (ME), which we discuss further below (Halberda, 2003; Markman, 1990; Mather & Plunkett, 2011; inter alia).

However, in most experiments of infant word comprehension, the named image is either unambiguously present or absent: while this is informative about the *limits* of children's knowledge, it does not reveal their finer-grained semantic representations. That is, if an infant looks at a cup upon hearing "cup", and looks away from the cup upon hearing "nose", this provides some evidence of a word-referent link in the child's lexicon. But it does not reveal whether the child's mental "cup" category also includes inappropriate category members (e.g., spoons or bottles). Understanding the referential specificity of infants' early words is germane to building a theory of early semantic development. Indeed, learning to properly extend the words in their lexicon is a many-years process; here, we examine some of its earliest stages.

18–24-month-olds not only recognize a large number of common words, but they are also aware of the relatedness among words. For instance, Luche, Durrant, Floccia, and Plunkett (2014) find that 18-month-olds differentially listen to lists of words that are semantically coherent, as opposed to lists that are from mixed semantic categories, suggesting that even in relatively small vocabularies, there is organization along semantic dimensions. Relatedly, when learning new words, two-year olds quickly deduce visual similarity among new referents, and listen longer to perceptually related words than unrelated words (Wojcik & Saffran, 2013). In further relevant work, Arias-Trejo and Plunkett (2010) find that by 18–24 months infants reliably fixate a named target when the target and foil referent are visually or semantically related (or neither), but still struggle when the two presented images are both semantically and visually related (e.g. "shoe" and "boot"). Taken together, these previous findings suggest that toddlers are using the relation between the words they know or have recently learned to guide their comprehension and lexical development, but may still struggle to differentiate between referents that share visual-semantic features.

The effects of semantic relatedness on visual fixations have also been examined in adults. Most germanely, Huettig and Altmann (2005) presented adults with two types of 4-picture displays: a

"target condition" in which a named target image appeared among distractors (e.g., piano, the named target, was displayed along with a goat, a hammer, and a carrot), and a "competitor condition" in which a semantic-competitor image (e.g., a trumpet) occurred among those same distractors, while the target (e.g., "piano") was said. (There was also a third condition with both a target and competitor displayed). The results indicated that target-looking was approximately 15% greater when the target was named with its correct label than with its semantically related label, while the overall number of saccades during the target word did not vary across these conditions (Huettig & Altmann, 2005). These results suggest that semantic relatedness drives eye movements toward related referents, even when a named target is not present.

In summary, previous research on everyday words shows that infants understand common nouns before their first birthday, and by their second birthday have a moderate level of understanding about how words are semantically related to each other. By adulthood, the networks among semantically related words are strong, and are measurable by eye movements during spoken word recognition.

Novel noun learning by inference

While infants get thousands of exposures in their first year of life to common nouns and their referents, learning contexts vary. The most transparent context is ostensive naming, where an infants' attention is directed to a single named object (e.g., mother holding a hotdog and saying "Look at the hotdog"). Although ostensive naming may occur commonly in some cultures, especially in speech to young infants, it may be a special case rather than the norm: it requires the co-occurrence of the word and the referent, in a sparse visual and linguistic environment, in an unambiguous labeling context. A second—perhaps more common—situation that can facilitate word-referent mapping, is hearing new words in the presence of familiar and unfamiliar objects. This is the classic case of mutual exclusivity (ME), whereby children infer that a new word labels a new object (Diesendruck & Markson, 2001; Horst & Samuelson, 2008; Markman, 1990, 1994; Mervis, Golinkoff, & Bertrand, 1994).

ME's underlying mechanism is a source of debate. While some accounts argue that recalling familiar objects' names is the critical step (Diesendruck & Markson, 2001; Markman, 1990; Mervis et al., 1994), others argue that the novelty of the new object and label are what drives learning and attention (Mather & Plunkett, 2010; Merriman & Schuster, 1991).

A further debate concerns *when* infants are first able to deduce that a newly heard label applies to a new referent. This is typically tested in the context of a two-object display where one object is familiar and "name-known" and the other is novel and "name-unknown." In this context, infants tend to exhibit strong baseline looking preferences before hearing any words; whether the preference is for the familiar or novel object seems to vary across stimuli and age (Bion, Borovsky, & Fernald, 2013; Halberda, 2003; Mather & Plunkett, 2010; White & Morgan, 2008). The generally accepted onset of ME behavior is in the middle of the second postnatal year (e.g., Halberda, 2003), with some evidence for novel labels drawing infants' attention at 10 months (Mather & Plunkett, 2010; cf. Weatherhead & White, 2016 for recent data with 10–12-month-olds in the phonetic domain.)

In related work with two-year-olds, Swingley and Fernald (2002) investigated word comprehension for familiar and unfamiliar spoken words. They found different fixation patterns as a function of word familiarity. When toddlers heard a *familiar* word that did not match the familiar image they were fixating, they rapidly looked away, searching for the target image. When they heard an *unfamiliar* word that didn't match the fixated familiar image (e.g., "meb"), their looks away were slower and more irregular. These results led the authors to suggest that hearing a word leads children to search their lexicon, rather than simply assess whether the heard word matches the visible image.

Thus, spoken word comprehension—as indexed by eye movements to named targets—varies both as a function of what the visual alternatives are, and how well the spoken label fits with a given visual

referent. Here we test how semantic specificity (or, put otherwise, word-to-referent "fit") varies in the context of two familiar objects, and in the context of one familiar and one unfamiliar object.

Under novelty-driven ME accounts, which hinge on attention to the novel word and label, infants' semantic specificity for the familiar object would be less relevant (although the relative novelty of the familiar object may play a moderating role). Under familiar-word driven ME accounts, in contrast, infants may be less likely to infer that a novel word refers to a novel object if they have uncertainty about the label of the familiar object. In this case, if a child knows that "spoon" is the label for spoon, and hears "dax" while seeing a spoon and a novel object, ME would help her learn the new word. But if she's not sure what the right label for a spoon is, and hears "dax", she may be less sure which image is being referred to. That is, semantic relatedness may modulate the applicability of the ME constraint on word learning. A first step in testing the hypothesis that there is an interaction between ME and semantic relatedness is to conduct the classic ME task (i.e., one known vs. one unknown and unrelated object) and ask whether performance on that task is correlated with infants' recognition of semantically related words. This manipulation is part of the present design.

Present study

In the present study, our primary question concerns infants' semantic representations of familiar words. Here we predict that if semantic specificity follows the same timeline as phonetic specificity, that by 12 months infants will look more at familiar objects when hearing their proper "matching" labels than when hearing related but "non-matching" ones. If semantic specificity is delayed vis-à-vis phonetic specificity, we would expect this pattern only in older infants. Because there is no evidence that suggests semantic specificity would precede phonetic specificity, we test infants 12 months and older.

Our secondary question probes comprehension in the context of a novel object, in an effort to bridge the novel word learning and familiar word comprehension literatures. By hypothesis, if infants anchor new word learning with known word knowledge, we would expect to see stronger novel word learning in children with more adult-like specificity in their word-form to semantic-category links. By testing infants up to 20 months, we include the standardly reported onset of ME (around 17 months, e.g., Halberda, 2003), and can probe whether comprehension of familiar words is linked to ME behavior during novel word learning.

To operationalize these questions, we presented pairs of images to infants: either both images were familiar common objects, or one was familiar and one was novel. We then presented infants with auditory stimuli directing them to look at (a) one of the familiar objects (*matching* trials), (b) an object that was *related* to one of the familiar objects (but was not actually present; *related* trials), or (c) the novel object, labeled by a nonce word (*nonce* trials).

The *matching* trials allowed us to assess infants' comprehension of familiar words in the context of familiar objects (Fernald et al., 1998), as a baseline. The *related* trials provided a measure of semantic specificity (i.e., does an auditory word-form trigger looking to semantically related referents to the same degree that it triggers looking to the matching referent?). The *nonce* trials provide a measure of ME (i.e., can infants infer that a new label refers to a novel object in the presence of a known object).

To be clear, we investigate semantic specificity here by assessing whether, given an instance of an object, an appropriate *matching* label leads infants to look at it more than an inappropriate but *related* label. This is not the same as asking whether infants know the difference between the meanings of two words pitted against each other directly. The head-to-head comparison of related items adds the burden of visual discrimination and feature overlap, testing knowledge of two semantically-similar (and thus generally visually-similar) words and two visual referents. Indeed as Arias-Trejo & Plunkett (Arias-Trejo & Plunkett, 2010) demonstrated, even 18–24-month-olds show poorer performance with semantically similar competitors, especially those that are also perceptually similar. As instantiated here, all trials featured two images that were highly discriminable and
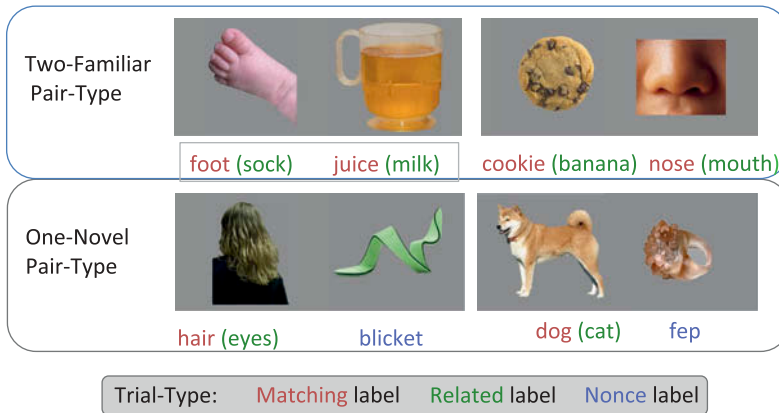
Figure 1. Experimental design. Infants saw each of the pairs of images shown above, 8 times each (with side counterbalanced across trials and orders). On matching trials (*n* = 12, 2 per image-pair) they heard the labels in pink, on related trials (*n* = 12, 2 per image-pair) they heard the labels in green (whose images never appeared), and on nonce trials (*n* = 8, 4 per nonce object) they heard the labels in blue. There were 16 trials in each pair-type, 32 trials total.

semantically unrelated. On matching and related trials, we test knowledge of how well each image fits with two different auditory labels. See Figure 1.

Methods

Participants

Participants were monolingual English-hearing infants from 12–20 months (*n* = 108, 51 female, at least 85% English spoken at home). Infants were recruited from the Rochester area through mailings, fliers, and phone calls, and had no reported hearing or vision problems. Families were compensated with $10 or a toy for their participation. Infants were recruited such that ~36 participated in the experiment in each of three age-groups: 12–14 months (*n* = 37), 15–17 months (36), and 18–20 months (35). 19 infants were excluded for fussiness or an inability to calibrate the eyetracker, leading them to contribute data to zero test trials (seven 12–14-month-olds, eight 15–17-month-olds, and four 18–20-month-olds). Forty-three further infants (fourteen 12–14-month-olds, fifteen 15–17-month-olds, and fourteen 18–20-month-olds) were excluded due to insufficient data contribution (i.e., failure to contribute data to at least 50% of the 32 trials: see Eyetracking Data Preparation below), leaving 46 infants in the final sample. See Table 1 for vocabulary, age range, and n of each group.[1]

Families who chose to complete demographics questionnaires (94%) reported that infants came from largely upper middle class homes: mothers worked on average 25.4 hours a week; 80% fell in

Table 1. Age and vocabulary characteristics of the sample. (Prod. Vocab. and Comp. Vocab refer to production and comprehension vocabulary on the Macarthur-Bates Communication Development Inventory, respectively).

Age-Group	Mean Age (mo.)	Mean Prod. Vocab.	Mean Comp. Vocab.	# Infants
12–14	13.55	8.93	81.53	16
15–17	16.11	22.92	116.00	13
18–20	19.61	89.88	136.00	17

[1]Our data inclusion criteria are delineated in "Eyetracking Data Preparation" below, but it should be noted that we experienced above-average data loss with the Tobii T60XL system. We have elected to retain only infants who contributed sufficient looking in a given trial (>25% of the 367–4000 msec analysis window), and in a given proportion of overall trials (50% or more) for this reason.

the three highest education categories of a 12-point scale, having attained a bachelor's degree or higher.

For the study itself, parents and infants were escorted into a testing room, and seated in a curtained, sound-attenuated booth, in front of a Tobii T60XL Eyetracker, running Tobii Studio (version 3.1.6). The experimenter, seated just outside the room, ran the calibration routine for the eyetracker, and then presented the infant with four "warm-up" trials, in which one image was shown on the screen, while a labeling sentence was played over the speakers in a child-friendly manner, e.g. "Look at the spoon!" (Fennell & Waxman, 2010). Parents were then provided with a visor or mask so they could not see the screen (confirmed by the experimenter on a closed-circuit camera throughout the experiment).

Once the experiment proper began, infants were shown two images side-by-side on the screen, and heard a sentence labeling one of 14 target words (12 common nouns and 2 nonce words.) Trials were 10 sec long, with the onset of the target word varying slightly for each trial; target onset was on average 3372 msec after the display of the images. Colorful attention getters accompanied by bird whistling noises occurred every eight trials to maintain infant interest.

After the eyetracking study, parents were asked whether they thought their infant might name some of the images. These infants ($n = 25$) participated in a task where the experimental stimuli were shown one at a time on an iPad. The infant was encouraged to label them by the experimenter or parent (with no phonetic or semantic cues provided). The timing of presentation of each image varied from 2–15 sec, based on the child's interest in the task; most children did not provide labels for any images (see "Results").

Design

Infants saw 32 test trials, lasting approximately 5 min total. On each trial, infants saw pairs of images while hearing a sentence that directed them to look at a noun. Each trial fell into one of two pair-types (*two-familiar* or *one-novel*) and one of three trial-types defined by which word was spoken (*matching, related, nonce*); see Figure 1. The *two-familiar* pair-type featured two familiar, common objects (foot-juice or cookie-nose). The *one-novel* pair-type featured one familiar, common object and one novel object (hair-blicket or dog-fep). On *matching* trials infants heard a sentence that labeled one of the familiar images on the screen (e.g., "Look at the foot!" while seeing the foot-juice image pair). On *related* trials infants heard a word that was semantically related to one of the images they saw, but did not actually appear (e.g., "Look at the sock!" while seeing the foot-juice image pair). On *nonce* trials, infants heard a novel word while seeing one of the one-novel pairs (e.g., "Look at the blicket!" while seeing a novel object and a hair image). There were *matching* and *related* trials for both pair-types, but *nonce* trials only occurred in the *one-novel* pair-type.

Trial order was pseudo-randomized into two lists that insured no image-pair (and thus no targets) occurred in back-to-back trials, and that the target image did not occur more than twice in a row on the same side of the screen. For each trial order, infants heard either the matching name or the related name for a given image for the first half of the experiment, and the other name for the second half of the experiment. For instance, in Order 1 infants heard "juice" and "sock" sentences when seeing the juice-foot pair, and in the second half they heard "milk" and "foot" sentences for that same image pair; Order 2 infants received the opposite name-image pairings in each half.

Materials

Visual stimuli

Infants saw images of twelve common nouns over the course of the experiment: four during the warm-up phase, and eight during the experiment. These images were photographs edited onto a plain grey background. During warm-up trials, a single image appeared centered on the screen (1920 x 1200 pixels); warm-up stimuli were apple, bottle, hand, and spoon. During test trials the

two images in a given pair (see "Design") appeared side by side, each taking up approximately half of the screen. Test images appeared on the left and right equally often across trials and orders, and included images of foot, juice, cookie, nose, hair, dog, blicket, and fep referents; see Figure 1.

Audio stimuli

Audio stimuli were 18 sentences containing a concrete noun. Four were used for warm-up trials (apple, bottle, hand, and spoon) and 14 for test trials. All sentences were prerecorded in infant-directed speech by a staff member with the local dialect, and normalized to 69 dB. Sentences were recorded in one of four carrier phrases: "Can you find the X?", "Look at the X!", "Where's the X?", or "Do you see the X?" (X represents the target word). Each image-pair used the same carrier phrase for all audio targets.

The two novel words were phonologically licit novel words ("blicket" and "fep"), while the matching and related audio targets were common nouns; see Figure 1. For the younger two age-groups, mothers heard the test sentences through over-ear headphones, and then repeated them aloud to their infant when prompted by a beep (Bergelson & Swingley, 2012; Shipley, Smith, & Gleitman, 1969). For the oldest group, all sentences were played over the computer speakers, because piloting had indicated discomfort and fussiness from children this age when their mother wore both a visor or mask (which was required so they could not see the screen), and headphones. There were also two attention-getting sounds in each trial: a pop when the images appeared on the screen, and a beep at the end of the pre-recorded sentence. These served to maintain infants' interest towards the screen.

Item selection

The nouns tested in this study were selected by finding a set of items that were approximately matched on a few key features. On average, matching items and related items each occurred over 200 times and were said by 14/16 mothers in the Brent corpus (Brent & Siskind, 2001). Both sets of words were reportedly understood by at least 70% of 12–18-month-olds in the English Words & Gestures section of the WordBank MCDI database (Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, 2016). Finally, all items had been used in looking-while-listening paradigms with infants from 6 months to 2 years of age (Bergelson & Swingley, 2012; 2013b, 2015; Fernald et al., 1998; *inter alia*).

The selected items were combined into visually presented pairs (cookie-nose, foot-juice, hair-blicket, dog-fep) that were perceptually and semantically unrelated. The items were further combined into matching and related dyads (cookie/banana, nose/mouth, foot/sock, juice/milk, hair/eyes, dog/cat) that did not share an onset phoneme, and critically, were semantically related. Given the limited vocabulary of one-year-olds, this resulted in a somewhat heterogeneous collection of nouns (see Figure 1). We return to this point in the discussion.

Procedure

Upon the family's arrival to the lab, staff explained the procedure to parents and received consent (approved by the University of Rochester IRB process). Depending on the mood of the child, parents either completed further paperwork before or after the eyetracking study, which included an optional demographics questionnaire, and one version of the Macarthur-Bates Communication Development Inventory (MCDI). Since a large proportion of the infants spanned the ages where both MCDI-Words and Gestures and MCDI-Words and Sentences are used (8–18 months and 16–30 months, respectively), parents of infants on the cusp were given the version of the form they deemed appropriate for their child's vocabulary. This resulted in 65% of families filling out the Words and Gestures form (Mean Age = 14.88, SD = 1.76), and 35% of families filling out the Words and Sentences form (Mean Age = 19.58, SD = 0.92).

Eyetracking data preparation

All eyetracking data were exported from Tobii Studio and analyzed in R.[2] All trials in which parents did not produce a target sentence, or produced the wrong word, were removed ($n = 17$ from $n = 9$ participants), along with lost trials due to technical error ($n = 6$, 1 participant). Then, all trials in which infants' eye gaze was not recorded for at least one fourth of the window of interest were dropped from analysis (due either to the child not looking at the screen, or the Tobii system's failure to record their gaze). Finally, participants were excluded from analysis altogether if they did not contribute the per-trial minimum data on at least 50% of trials, due to fussiness, data loss due to poor track from the Tobii system, complete inability of the system to calibrate and locate their pupil, or parental error or interference ($n = 62$). This procedure resulted in the retention of data from approximately 3/4 of the data from each included participant for each trial-type ($M_{matching}$ = 8.74 (1.96) out of 12 trials, $M_{related}$ = 8.76 (1.91) out of 12 trials, M_{nonce} = 5.85 (1.52) out of 8 trials).

Results

Given the inherent distributional properties and level of noise in infant eyetracking data, we adopted a conservative analysis approach. We measured target looking within a post-target window from 367–4000 msec after target onset (Fernald et al., 1998; Swingley & Aslin, 2000). We provide the smoothed timecourse of our data for visualization purposes, with vertical lines delineating our analysis window in the Appendix.

We calculated baseline-corrected proportion of target looking (propTcorr) by subtracting the proportion of time that infants looked at a given image before the target word was spoken (baseline window) from the proportion of looking during the analysis window (i.e., (baseline_T/ (baseline_T + baseline_D))—(T/(T + D), T = Target, D = Distractor)). This measure adjusts for baseline preferences that infants may have for fixating images in the absence of a spoken label, and then expresses their change in preference from this baseline after the target word is spoken. Thus, this measure is compared to a chance-level of 0 and ranges from −1 to 1.[3]

We analyze the data using mixed effects models, and simple significance tests, using propTcorr as our outcome measure. We first report modeling results for matching and related trial-types (i.e., without the nonce trials) because that was our central question of interest.

While certain subsets of the data did not differ from a normal distribution, given that these data are generated by differences among proportions, we rely on model comparison based on ANOVA results, and use non-parametric follow-up tests. All tests below are two-tailed; this is a more conservative measure of significance given that there are clear directional predictions for each test we conducted. We begin with the analyses for our key question of interest: do infants show differentiated comprehension when they hear a related label as opposed to a matching label, and does age influence this differentiation? Thereafter, we address our second question: how does novel word comprehension (in a mutual exclusivity setting) compare to comprehension of matching and related labels? We then examine results by trial-type and pair-type, production results, and vocabulary measures.

GLMM Results for matching and related trials

We fit Generalized Linear Mixed Effects Model (GLMMs; Baayen, 2008) using R's lme4 and lmerTest packages. Beginning with a baseline model with random effects for subjects and image-pairs, we examined effects of trial-type (matching, related), pair-type (two-familiar, one-novel), and age (in months, mean-centered). The results of the model comparison revealed that the best model

---

[2]All code generating the manuscript and results in this article, along with the data and its analysis pipeline, are available through the first author's website.

[3]Using baseline-corrected target looking is the standard measure in the field, given infants' large saliency biases for certain images, which are generally orthogonal from the question under investigation and lead to idiosyncratic item effects. However, this corrected measure too is imperfect. We provide the corrected and uncorrected data for our readers in Figures 6 and 7

includes an trial-type x age interaction (but not pair-type, either additively, or in an interaction term). See Table 2 for ANOVA results.

To explore the interaction between trial-type and age we computed *semantic advantage scores* by subtracting proptTcorr in all related trials (e.g., sock called "foot") from propTcorr in all matching trials (e.g., foot called "foot"), i.e., matching propTcorr—related propTcorr, collapsing over pair-type. There were six such matching–related dyads (foot/sock, juice/milk, cookie/banana, nose/mouth, dog/cat, and hair/eyes). Critically, this measure was computed on a per-child basis using the baseline-corrected measures generated from the trial-level data. This measure provides an index of semantic specificity: positive scores indicate more target-looking in the matching trials than in the related trials, negative scores vice versa.

Figure 2 shows each infant's semantic advantage scores as a function of age. As suggested by the interaction in our GLMM, age and semantic advantage scores are correlated: older infants have stronger semantic advantage scores than younger infants (Kendall's τ = 0.204, p = 0.046.) That is, older infants look more at images when hearing a matching label than when hearing a related label.

GLMM results across all trial-types

We next turn to overall analyses to ascertain whether infants (a) provide evidence for understanding the familiar words, (b) perform differently when shown two-familiar vs. one-novel displays, and (c) look more at the novel object upon hearing the nonce word.

Table 2. Test of fixed effects from final GLMM, excluding "nonce" trials. Note: degrees of freedom estimated using the Satterthwaite approximation. Random effects for participants and image-pairs were also included in this model. Num. and Den. refer to numerator and denominator, respectively.

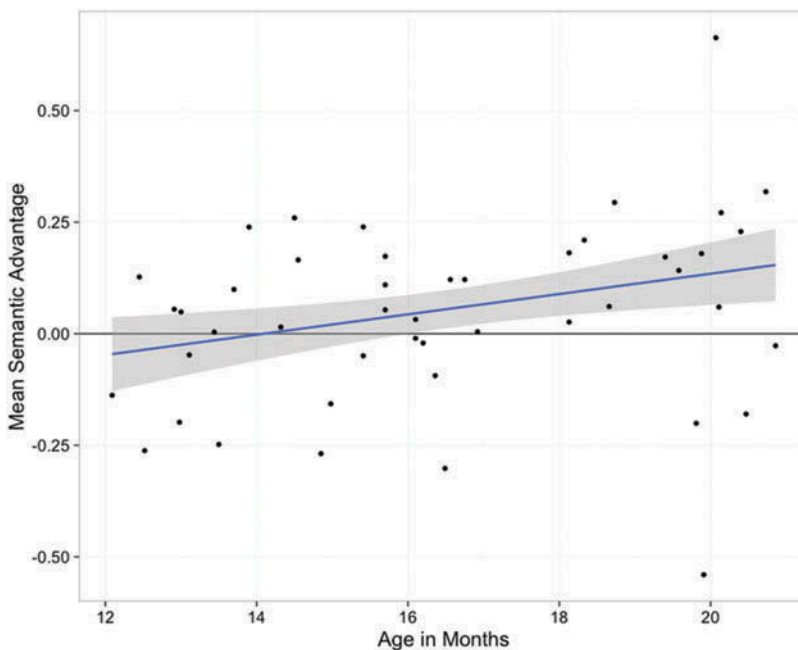| Param.	SS	Num. DF	Den. DF	F	p-value
AgeMonthsCntr	0.018	1	46.563	0.143	0.707
TrialType	0.184	1	727.152	1.458	0.228
AgeMonthsCntr:TrialType	0.535	1	724.660	4.237	0.040

Figure 2. Semantic advantage scores by age. This figure shows how much more each infant looked at the target image on matching vs. related trials. The fitted line is a robust linear fit with 95% confidence intervals in grey.

We fit overall GLMMs with random effects for subjects and image-pairs, and examined the effects of trial-type (matching, related, nonce), pair-type (two-familiar, one-novel), and age (in months). Nested model comparison indicated improved fit when trial-type and pair-type were included (by both AIC and LRT; all $p < .05$ by Chi-Square test for LRT analysis), while (mean-centered) age did not significantly improve model fit (when added to the baseline model, or to a model with trial-type, pair-type, or both; all $p > .05$ by Chi-Square test for LRT analysis).[4] A pair-type x trial-type interaction was not justified by nested model comparison, and moreover, was rank-deficient given no "nonce" trials in the two-familiar pair-type ($p > .05$ by Chi-Square test for LRT analysis); see Table 3 for test of fixed effects in the final overall GLMM.

In the next three sections, we probe the effects of trial-type and pair-type, and we also provide results by age-bin in order to facilitate comparisons with previous work, which uses smaller age-ranges. Most germanely, (1) Swingley and Aslin (2002), and Mani and Plunkett (2010) found phonetic specificity effects at 12–14 months, and (2) Halberda (2003) found different ME effects in 14-, 16-, and 17-month-olds. Given that age was only a significant predictor in our modeling results when holding out the nonce trials, but not in our overall GLMM, we do not conduct pair-wise age-bin comparisons (e.g., 12–14-month olds vs. 15–17-month-olds, etc.).

Results by trial-type

Descriptively, infants' target-looking on nonce and matching trials increased over the 12–20 month age range, while their target-looking when hearing semantically related labels did not; see Figure 3. At the whole group level, we find strong performance for matching and nonce trials (significant by binomial and Wilcoxon tests); however, performance on related trials was only marginally above chance (see Table 4). That is, overall, infants looked at the named target image significantly more in the analysis window than in the baseline window when hearing an appropriate label (be it a familiar word or nonce word), but only marginally so when hearing a word that is semantically related to the target image; see Figure 3 for subject-level data for each trial-type and Figure 6 for all baseline, analysis window, and propTcorr data by item.

Results for two-familiar pair-type

Unsurprisingly, infants performed well on matching trials in the two-familiar pair-type, i.e., when hearing a matching label for one of two familiar images. Indeed, infants in each age-group attained positive item-means on matching trials for each of the four items. Over subjects, each age-group's propTcorr was significantly above chance; see Table 5. These results confirm that our procedures were reliable in replicating earlier studies at these ages.

On related trials, performance was lower on average than it was for matching trials, for each age-group. However, for the two younger groups, performance was quite similar on related trials and matching trials: in the two-familiar pair-type, 12–14- and 15–17-month-olds' looked more at the target image when they heard the related label, for all four items, just as they had

Table 3. Test of fixed effects from final overall GLMM. Note: degrees of freedom estimated using the Satterthwaite approximation. Random effects for participants and image-pairs were also included in this model. Num. and Den. refer to numerator and denominator, respectively.

Param.	SS	Num. DF	Den. DF	F	p-value
TrialType	1.575	2	981.779	5.984	0.003
PairType	2.231	1	4.117	16.957	0.014

[4]Including age as our 3-month age bins rather than as a continuous variable rendered the same pattern of results.
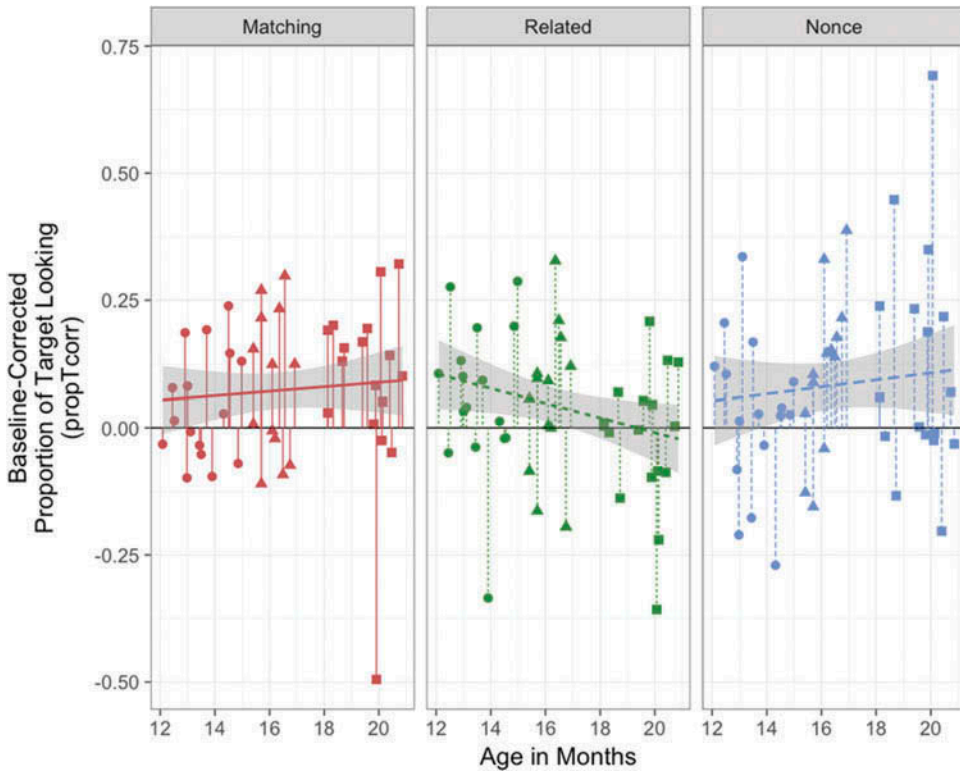
Figure 3. Subject mean performance across trial-type. Each point indicates a given infants' subject mean for a given trial-type (indicated by panel and color). The shape of the point indicates each infant's age-group. The fitted line is a robust linear fit with 95% confidence intervals in grey. Overall performance on nonce and matching trials was significantly above chance; see Table 4.

Table 4. Overall performance by trial-type. Mean, SD, and p-values calculated over subjects in each age-group

Trial-Type	#Infants with pos. means by trial-type (binomial test p-value)	Mean (SD) propTcorr	Wilcoxon Test *p*-value
Matching	31/46 (p = 0.026)	0.073 (0.147)	0.000
Nonce	31/46 (p = 0.026)	0.084 (0.187)	0.005
Related	29/46 (p = 0.104)	0.031 (0.147)	0.075

when hearing the matching label. In contrast, the 18–20-month-olds showed numerically worse performance with related labels than matching labels, for each image; see Table 5 and Figures 4 and 6. This pattern suggests robust performance on our two-familiar matching-label trials, but more mixed performance across 12–20 months on the related trials, consistent with the age x trial-type interaction we found in our initial GLMM.

Table 5. Two-familiar pair-type performance by age-group and trial-type. Mean, SD, and p-values (Columns 5 and 6) are calculated over subject-means for each trial-type for each age-group (i.e., for each row).

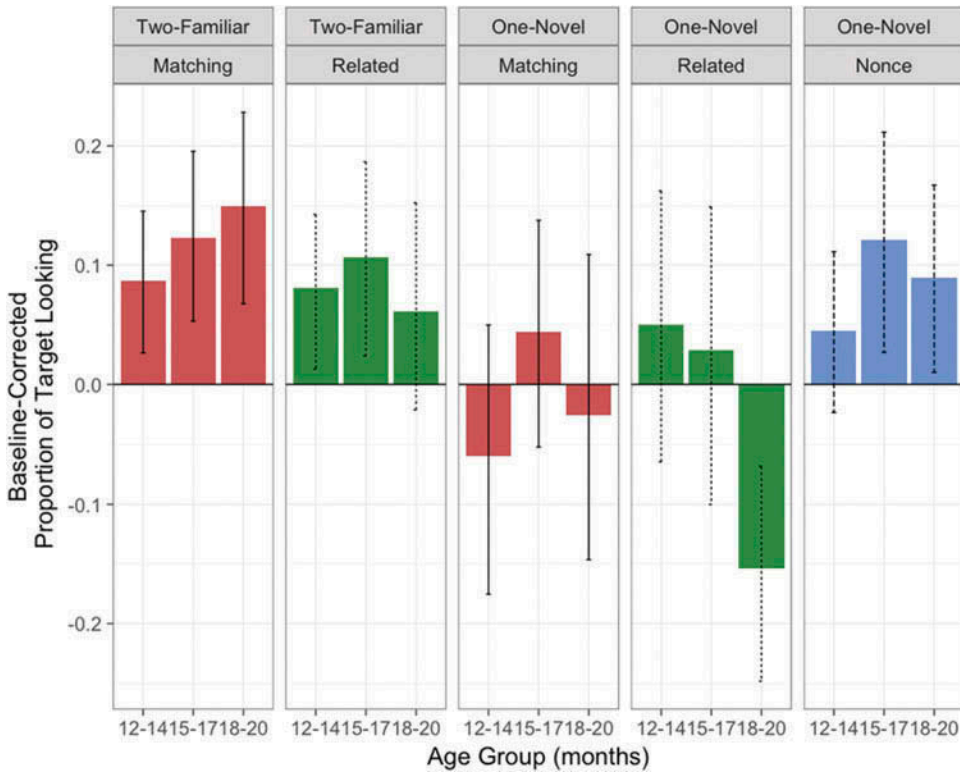| Trial-Type	Age-Group	Pos. Subj. Means	Pos. Item Means	Mean (SD) propTcorr	Wilcoxon Test *p*-value
Matching	12–14	11/16	4/4	0.09 (0.14)	0.039
Matching	15–17	10/13	4/4	0.11 (0.15)	0.040
Matching	18–20	13/17	4/4	0.16 (0.2)	0.013
Related	12–14	10/16	4/4	0.07 (0.16)	0.074
Related	15–17	10/13	4/4	0.09 (0.13)	0.027
Related	18–20	10/17	3/4	0.06 (0.22)	0.225

Figure 4. Mean performance by trial-type, pair-type, and age-group. Each bar represents aggregated performance across infants in each age-group. Error bars indicate non-parametric 95% confidence intervals around the mean; see text for details.

Results for one-novel pair-type

Due in part to large baseline biases in favor of the familiar images, described further below, no age-group showed significantly positive propTcorr values for related or matching trials for the one-novel pairs, dog-fep and hair-blicket (see Table 6). While nonce-trial performance over all infants was significant, this effect was clearer for older infants: indeed, while 12–14-month-olds were at chance across the two nonce items, 15–17-month-olds performance was significantly above chance, and 18–20-month-olds' performance was marginally significant. This is consistent with previous research on ME (e.g., Halberda, 2003); we return to this point in the discussion.

Table 6 One-novel pair-type performance by age-group. Mean, SD, and *p*-values (Columns 5 and 6) are calculated over subject-means for each trial-type for each age-group (i.e. for each row).

| Trial-Type	Age-Group	Pos. Subj. Means	Pos. Item Means	Mean (SD) propTcorr	Wilcoxon Test *p*-value
Matching	12-14	7/15	1/2	-0.06 (0.28)	0.679
Matching	15-17	7/13	1/2	0.03 (0.19)	0.588
Matching	18-20	7/17	1/2	-0.07 (0.34)	0.431
Related	12-14	8/15	1/2	0.03 (0.22)	0.599
Related	15-17	6/13	1/2	-0.04 (0.29)	0.787
Related	18-20	3/17	0/2	-0.16 (0.18)	0.002
Nonce	12-14	11/15	2/2	0.03 (0.16)	0.359
Nonce	15-17	10/13	2/2	0.11 (0.16)	0.040
Nonce	18-20	10/17	1/2	0.12 (0.23)	0.089
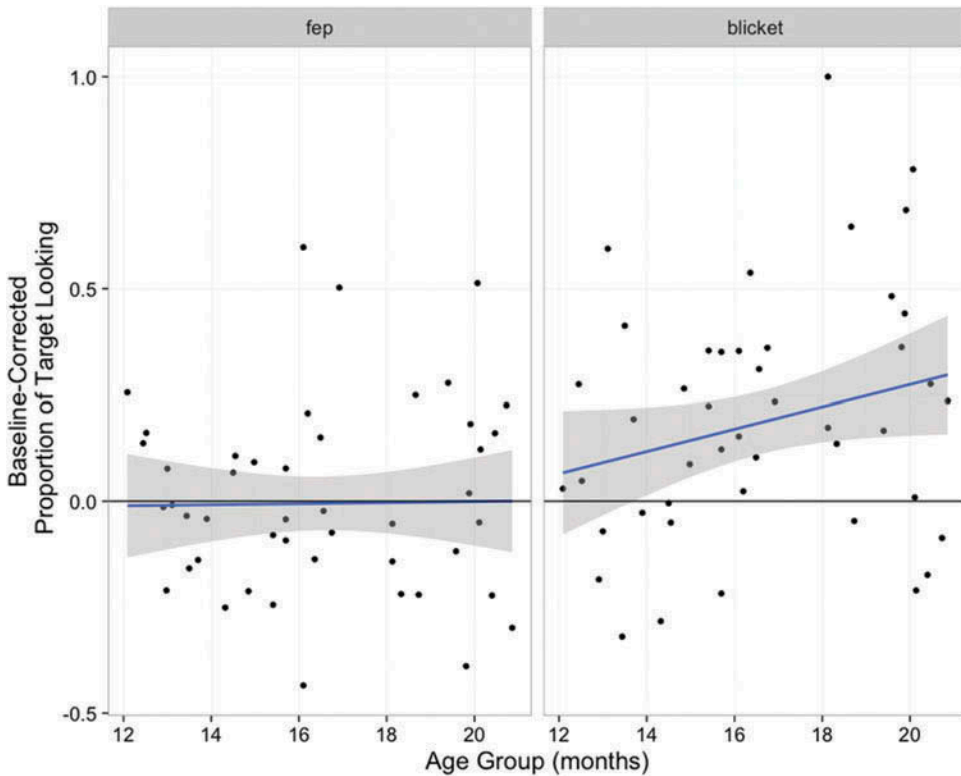
Figure 5. Nonce item means by subject.The fitted line is a robust linear fit with 95% confidence intervals in grey.

We provide descriptive results separately for each one-novel pair, given the large baseline bias. For the dog-fep pair, results followed the same pattern as the two-familiar pair-type described above: across the board, infants looked more at the dog after hearing "dog" than in the baseline window; the two younger groups also looked at the dog image more after hearing "cat" than in the baseline window, while the eldest group did not increase looking to the dog upon hearing "cat"; see Figure 6. Infants found the dog image particularly attention-garnering for this pair: across our sample, they looked at the nonce object only 35% of the time, in both the baseline and target window (i.e., after hearing it labeled "fep".)

The familiar-image baseline bias was particularly notable for the hair-blicket pair, rendering somewhat paradoxical results for matching and related trials. In these trial-types, infants looked at the hair image more during the baseline period than during the target window, whether they subsequently heard the matching label "hair" or the related label "eyes". In contrast to the dog-fep pair, infants in all three age-groups looked more at the novel object upon hearing the nonce word "blicket" (mean propTcorr for 18–20-month-olds = 0.22, 15–17-month-olds = 0.23, 12–14-month-olds = 0.08).

Given that we did not predict our two one-novel pairs to show such different patterns, we conducted a purely exploratory analysis to compare infants' performance on our two nonce items. We find that infants across our age-range were squarely at chance with "fep", but showed stronger and improving performance with "blicket"; see Figure 5.

In summary, our eyetracking results show that infants looked most reliably at the named image on the standard two-familiar matching trials. Performance on the one-novel trials was highly variable across our two item-pairs for both familiar and novel items; nonce item performance improved somewhat with age. Finally, across pair-types, performance on matching trials was better
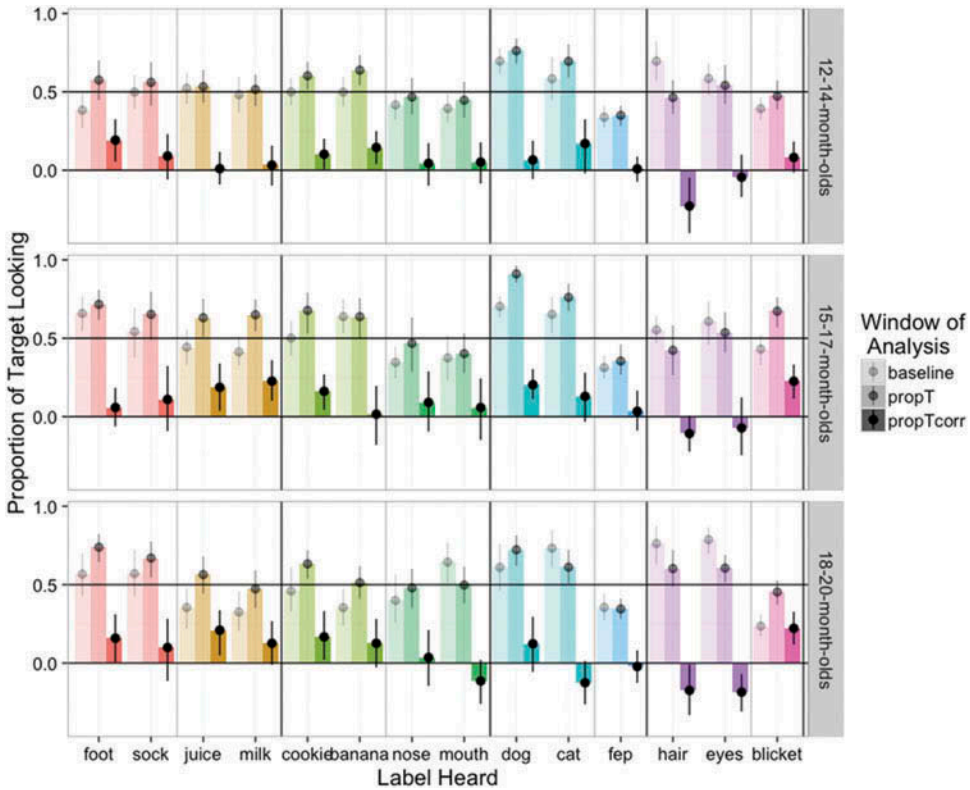
Figure 6. By-item performance across windows. Items are indicated across the x-axis. Colors denote each visual target; the left-side triplet of a given color is for the matching audio target, the right-side triplet for the related audio target. Dark black lines separate each image-pair, and grey vertical lines separate the two items within each pair. The y-axis indicates the proportion of target looking for each item by window of analysis (shading indicates the window: lightest bars indicate the baseline window, middle-shading bars indicate the target window, propT, and the darkest bars are the subtraction of the two: propTcorr). For instance, in the top left we see 18–20-month-olds' (top panel) proportion of looking at the foot image (all teal bars) on "foot" audio target trials (leftmost 3 bars) in the baseline, target, and target-corrected-for-baseline window. This figure is provided for data visualization transparency.

than on related trials as age increased from 12–20 months, suggesting increasing semantic specificity over the second year.

Production task results

Fifteen of the 25 infants who were asked to label the test images (based on their parent's indication of whether their child was producing such object labels reliably) provided at least one label, resulting in 67 productions. All but one infant who produced items was in our 18–20-month-old group (Mean Age = 19.38, 16.2–20.96). Infants produced words (with recognizable if not adult-like phonology) for 1–7 of our 8 target images (M = 4.47; SD = 2.42.) Of the 67 labels produced, 9 were attempts to label the nonce objects (none with correct phonetic content). Of the remaining 58 labeled images, 51.7% received the matching label, 6.9% received the related label, and 41.4% received another label not heard in the experiment (e.g., calling the "hair" image "gramma"). That is, even though all infants had just heard each of our familiar images labeled by matching and related nouns an equal number of times, infants predominantly provided the appropriate matching name or an unheard name, and only very rarely provided the semantically related name.

MCDI results

Both comprehension and production vocabulary as assessed by the MCDI were significantly correlated with age (comprehension vocabulary: Kendall's $\tau = 0.34$, $p = 0.01$; production vocabulary: Kendall's $\tau = 0.43$, $p < .001$). Further vocabulary analyses failed to reveal any additional results of interest and are omitted for brevity.

Discussion

This study examined word comprehension across several contexts. Consistent with previous research, we found that in the standard experimental display with two familiar nouns, one of which is named, infants from 12–20 months looked more at the named image. We then examined two further contexts of word recognition: (a) those in which infants heard a word-form that was semantically related to one of the images and (b) those in which infants heard a nonce word labeling a novel object.

For the first context, we found that younger infants looked at a visual referent (e.g., foot) to the same degree whether hearing a matching label ("foot") or a related label ("sock"). In contrast, older infants look more at the referent when hearing its matching label than when hearing its related label. These results suggest that infants can reliably pick out named referents among two options on a sparse display, but their eye movements are equally triggered by related words and matching words, with a gradual shift in favor of the matching word by around 1.5 years of age. Put simply: these data suggest that for familiar objects, infants know aspects of word meaning months before they show more adult-like semantically differentiated categories, as indexed by eye movements to a referent following a spoken word.

These results are consistent with a set of related accounts. For instance, it may be the case that for younger infants, the auditory word-form triggers a family of semantically related referents while for the older infants *only* the appropriate referent is triggered. Another possibility is that infants know these familiar words to the same degree across our age-range, but younger infants employ a "best fit" strategy more readily than older infants, not due to less lexical knowledge early on, but to a greater use of a "which picture matches best" heuristic. It could also be that for all infants, hearing the auditory word triggers a set of related words, but that for older infants the appropriate concept is *more strongly* activated. This last account seems to us more consistent with the toddler and adult literature, where participants' eyegaze is pulled to semantically or conceptually related referents in the context of an auditory word (Bergelson & Swingley, 2013b; Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005).

For our other context of word recognition, i.e., nonce word comprehension, we found that as a group, infants looked significantly more at the nonce object after hearing it named than they did in the baseline window, though this appears to be driven by one of our two nonce items in particular (see Figure 5.) While even younger infants may have rudimentary ME knowledge that our experiment and analysis was unable to capture (cf. Mather & Plunkett, 2010), our results are in keeping with previous literature placing the emergence of ME around 17 months (Halberda, 2003).[5] As our results also highlight, however, novel word learning may be strongly influenced by baseline preferences: further work is needed to uncover the relation between infants inherent biases to fixate familiar or novel objects, and their ability to link new words to new objects that enter the lexicon in a long-term way.

Our results leave open whether there is a causal relationship between semantic specificity and ME abilities. While we find improving performance on matching vs. related trials with age (i.e., greater semantic specificity; see Figure 2), and stronger ME skills in the two older groups than in the youngest group, this pattern has several interpretations. For instance, it may be that older infants are just better at all tasks, and thus their success in both of these word comprehension contexts does not

speak to a link between them. More enticingly, older infants' more adult-like semantic categories may be more deeply related to their ability to quickly infer the referent of novel words.

That is, increasing certainly about what a word's meaning is (or is not) may lead to a greater willingness to infer that new words map onto new objects. Such an account, compatible with the set of ME accounts that highlight the role of the familiar word (Markman, 1990), would suggest that younger infants have less "certainty" about what a spoken word means, and are thus less sure whether it maps onto a familiar object or a novel object. As their certainty increases with age, so too does their willingness to rule out a familiar object as the referent for a new name. This, in turn, may help explain why even the familiar items on one-novel trials were harder for infants than on the two-familiar trials: the novel object's presence may increase infants' uncertainty (or tax their attentional resources) across the board. Such an account too may speak to 18–20-month-olds' performance on the one-novel related trials (see Figure 4), where they looked significantly more at the *nonce* object upon hearing the *related* word, as if to say "that's surely not the familiar object's label, perhaps it labels the new object." A certainty-related account is compatible with recent toddler work as well, where children showed stronger ME behavior when they could *produce* the familiar object label than when they only understood it (Grassmann, Schulze, & Tomasello, 2015). While the infants in the present work have quite small productive vocabularies, a similar notion of "certainty" may apply to the present context of word comprehension.

Thus, both a global-improvement and a word-learning specific account offer possible explanations consistent with the present findings and with prior data. Indeed, the large baseline effects in the nonce-item pairs in the present dataset preclude a confident comparison across the nonce trial and semantic advantage score measures. As instantiated here, there is no correlation between them (Kendall's $\tau = -0.13$, p = 0.199). To test this proposal directly, future research must focus on 1.5–2 year olds' semantic specificity over a larger set of items in conjunction with querying their ME abilities over a broader set of novel objects. It is also important that such further work involve multiple tokens per type, as is the general case in newly acquired words and categories but rarely the case for in-lab studies.

Unsurprisingly, infant word comprehension improves with age. How to quantify this improvement is less straightforward. While some in-lab measures show steady improvement in response-time and accuracy to familiar words, or growing lexicon size, (Dale & Fenson, 1996; Fernald et al., 1998), not all in-lab measures reflect this improvement, especially when large continuous age ranges are tested (see Bergelson & Swingley, 2012, pp. 6–13-month-olds). In the present data, we found clearest evidence for an age effect in the interaction between trial-type and age on non-nonce trials (confirmed by significant interaction in our model and in the correlation between age and our semantic advantage scores). Given that we found several other age-related trends (not all of which reached statistical significance), we now turn to a discussion of the role of age in our results.[6]

First, there were likely large individual differences in the present dataset both due to infants' knowledge, and due to the relatively high rate of data loss from the Tobii eyetracker. This would increase within-group variability, potentially swamping a smaller overall age effect. Second, while we expected monotonic improvement with age on nonce and matching trials, the predictions for the related trials were less clear based on previous research. For related trials, the adult-like behavior is to show *diminished* performance relative to the matching trials. We limit speculation about the underlying developmental trajectory pending further research but find it worthwhile to note that in our view, isolated experiments testing a single age are less likely to inform our theory of this developmental path than cross-age approaches that use the same stimuli over a continuous swathe of development. The default assumption in developmental research is that children get better over time: when investigating an ability with the opposite prediction, denser and wider sampling will likely prove more fruitful.

---

[5]Indeed we find modest support in our own data for the kind of discontinuous ME emergence Halberda suggests: while age did not correlate with "blicket' comprehension when age was computed as a continuous variable, age-group (12–14, 15–17, 18–20 months) was significantly correlated with "blicket' comprehension, i.e. propTcorr (Kendall's *tau* = 0.25, *p* = 0.036).

Whatever the relationship between name-known word recognition and new word learning with age, infants must properly constrain both the sounds of a given word and its meaning in order to acquire the lexicon of their community. The present data suggest that some constraints on the semantics of familiar words do not appear until around the midpoint of the second postnatal year. This result is especially intriguing in light of two previous sets of findings. First, previous research suggests that infants begin to understand familiar words in contexts similar to the present study around 6 months of age (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012); here we find that even a year later infants continue to acquire greater precision in their word meanings, as operationalized in the current research. This suggests that there may be a protracted "semantic narrowing" phase between ~6–18 months (or, behaviorally indistinguishably, a decreasing reliance on a "best fit" heuristic). While further research is needed to uncover what exactly is changing in infants' word-meanings over this time, here we suggest that infants' initial semantic representations are overly broad. This leaves open whether infants have one large category for, e.g., feet, socks, and feet with socks on them, or whether a given single category is overly broad. A more psychophysical approach may help to pull these possibilities apart, though infants' patience in providing such data in a single testing session, requiring many repetitions of the same stimuli, may be limited.

Second, the present results suggest that semantic precision for familiar nouns comes online later than phonetic precision. That is, previous research has indicated that the phonetic representation of common nouns is adult-like by 12–14 months of age (Mani & Plunkett, 2010; Swingley & Aslin, 2002), while here we see analogous semantic precision emerging closer to the middle of the second year. This pattern of data is in keeping with an overlapping cascade in early acquisition, where infants' native language phonology becomes solidified before their semantics. One potential direction for future work is to manipulate both semantic and phonetic precision within the same experiment, to get within-child measures of these kinds of variability, and their interrelatedness. This may be an especially fruitful avenue for further research given that infants must deduce that similarity in sound space does not generally result in similarity in semantic space (holding aside the special case of sound-symbolism, cf. Bankieris & Simner, 2015).

Unlike phonetic space, where the features of interest are generally well agreed upon to a first approximation (e.g., voicing, rounding, etc.), in semantic space the features themselves prove elusive. Indeed many years of debate in the philosophy of language and concepts literature have highlighted how difficult it is to operationalize meaning or break it into agreed upon component parts or dimensions (Casasanto & Lupyan, 2015; Fodor, Garrett, Walker, & Parkes, 1980; Katz & Fodor, 1963).

Our familiar noun items were selected to optimize several parameters when considering the words overall, the two words that appeared together visually (e.g., foot-juice), and the related-matching audio word dyads (e.g., foot/sock). We picked items and pairs with an eye toward controlling words' frequency in the input to children based on published corpora, and high rates of word knowledge based on parental report on MCDI measures across the 12–20 month age range (Brent & Siskind, 2001; Fenson et al., 1994). This resulted in a somewhat heterogeneous set of related-matching contrasts. That is, while "nose" and "mouth" are both body parts and "cat" and "dog" are both animals, there is more visual feature overlap in the latter case (four-legged, has tail, etc.), but likely more opportunity for pointed contrast in the former (i.e., all babies have noses and mouths, but not all babies' households have cats and dogs).[7]

Relatedly, the precise stimuli we used resulted in fairly large baseline-differences in looking time to our pairs that included nonce objects, just as has been found in previous studies (Halberda, 2003; Mather & Plunkett, 2011). While this seems somewhat unavoidable given the constraints inherent in

---

[6]Only the oldest infants heard the test sentences over computer speakers, while the younger two groups heard them delivered by their parent live. It is in principle conceivable that this presentation difference would lead to differential performance only on the related and nonce trials (since all age-groups showed above-chance performance on matching trials). We find such an account unlikely given that (1) it is unclear why this would only matter for a subset of trials, and (2) there is a body of previous research showing one- and two-year-olds' comprehension of spoken words with both parent and speaker-produced stimuli (see Bergelson & Swingley, 2012; Bergelson, Shvartsman, & Idsardi, 2013; Fernald et al., 1998, 2008; *inter alia*; Tincoff & Jusczyk, 1999, 2012)

stimulus selection for infants, further research is necessary to ensure that our results hold across a wider range of nonce and familiar words and their referents.

In the present paradigm, eye-movements are a stand-in for infants' underlying lexical knowledge. Encouragingly, the pattern of results we find in production matches what we find in comprehension: while by and large only the eldest infants provided labels for our familiar images, when they did so they overwhelmingly provided either the matching name, or a name not used in the experiment (93.1%), rather than the related name (6.9%). This is true even though they just heard the matching and related names in the context of the same images an equal number of times. While younger infants did not show a semantic advantage for the matching over related name, it remains possible that even younger infants' knowledge is more refined than the eyetracking results suggest.

Anecdotally, one 15-month-old showed verbal differentiation on one set of matching and related trials. For the cookie-nose images: when hearing "Do you see the cookie?" the child answered "yes", but when hearing "Do you see the banana?", the related (but not present) prompt, the child said "no." While not enough children provided this kind of feedback to allow for analysis, this anecdote hints at a layer of knowledge perhaps uncaptured by our eyetracking measures. Indeed it is a bit odd to ask participants to look at something that is not present as we did here on related trials: the finding that adults look at the related image when more appropriate images are not present, as in the work of Huettig and Altmann (2005) discussed above, may reflect not just semantic specificity, but a theory of the task at hand. Infants too, it stands to reason, bring to the experiment their own lexical knowledge, and their understanding about ambiguity and communicative intent, more broadly construed.

In summary, the present set of results suggests that semantic specificity emerges over a rather lengthy period in the second postnatal year, during which infants' understanding of familiar words becomes more firmly constrained. We find that from 12–20 months infants readily recognize familiar words in the context of familiar objects. During this same time, the presence of novel objects seems to complicate spoken word comprehension, leading to evidence of ME in older infants. Furthermore, across this age-range, infants grow to preferentially respond to matching word-image pairs over semantically related word-image pairs. Notably, this sensitivity to semantic fit emerges months later than analogous responses for non-prototypical *phonetic* content in familiar words, as found in previous research (e.g., Swingley & Aslin, 2002).

In conclusion, the present research provides two key contributions to the literature. First, it suggests that phonetic specificity emerges earlier and more rapidly than semantic specificity. It also suggest that one underappreciated component of early lexicon building may entail not just learning what words refer to in the world, but honing category boundaries among related concepts as well.

ORCID

Elika Bergelson 🆔 http://orcid.org/0000-0003-2742-4797

---

[7]An analysis using word2vec over the CHILDES corpus shows that our matching-related pairs were each more related (quantified as cosine similarity) than the related word was to the other displayed image; code available on github: [https://github.com/SeedlingsBabylab/w2v_cosines/]

Richard Aslin ⬤ http://orcid.org/0000-0003-4092-1712

References

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1–2), 63–80. doi:10.1016/j.jecp.2009.10.002

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Retrieved from https://books.google.com/books?hl=en{\&}lr={\&}id=UvWkIg5E4foC{\&}oi=fnd{\&}pg=PA335{\&}dq=Baayen,+R.+H.+(2008).+Analyzing+linguistic+data:+A+practical+introduction+to+statistics+using+R.+Cambridge+University+Press.{\&}ots=1YHlfjo3Hh{\&}sig=lZzOGCMMiq7KgCUnGKdRhJSLWHI

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, *136*, 186–195. doi:10.1016/j.cognition.2014.11.013

Barrett, M. (1978). Lexical development and overextension in child language. *Journal of Child Language*, 205–219. Retrieved from http://journals.cambridge.org/abstract{_}S030500090000742X

Bergelson, E., Shvartsman, M., & Idsardi, W. J. (2013). Differences in mismatch responses to vowels and musical intervals: MEG evidence. *Plos One*, *8*(10), e76758. doi:10.1371/journal.pone.0076758

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. doi:10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013a). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397. doi:10.1016/j.cognition.2013.02.011

Bergelson, E., & Swingley, D. (2013b). Young toddlers' word comprehension is flexible and efficient. *Plos ONE*, *8*(8), e73359. doi:10.1371/journal.pone.0073359

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*(4), 369–380. doi:10.1080/15475441.2014.979387

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, *126*(1), 39–53. doi:10.1016/j.cognition.2012.08.008

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–598. doi:10.1111/desc.12242

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44. doi:10.1016/S0010-0277(01)00122-6

Casasanto, D., & Lupyan, G. (2015). All concepts are Ad Hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge, MA: MIT Press.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459. doi:10.3758/BF03193787

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, {&} Computers*, *28*(1), 125–127. doi:10.3758/BF03203646

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630–641. doi:10.1037/0012-1649.37.5.630

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*(5), 1376–1383. doi:10.1111/j.1467-8624.2010.01479.x

Fenson, L., Dale, P. S., Reznick, J. S., & Bates, E. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, 5. Retrieved from http://www.jstor.org/stable/10.2307/1166093

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*(3), 228–231. doi:10.1111/1467-9280.00044

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, 97–135. doi:10.1017/CBO9781107415324.004

Fodor, J. A., Garrett, M. F., Walker, E., & Parkes, C. H. (1980). Against definitions. *Cognition*, *8*(3), 263–367. doi:10.1016/0010-0277(80)90008-6

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 1–18. doi:10.1017/S0305000916000209

Gelman, S. A., Croft, W., Fu, P., Clausner, T., & Gottfried, G. (1998). Why is a pomegranate an apple? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of apple and dog. *Journal of Child Language*, *25*(2), 267–291. doi:10.1017/s0305000998003420

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1200. doi:10.3389/fpsyg.2015.01200

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–B34. doi:10.1016/S0010-0277(02)00186-5

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. doi:10.1080/15250000701795598

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–32. doi:10.1016/j.cognition.2004.10.003

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*(2), 170. doi:10.2307/411200

Luche, C. D., Durrant, S., Floccia, C., & Plunkett, K. (2014). Implicit meaning in 18-month-old toddlers. *Developmental Science*, 1–8. doi:10.1111/desc.12164

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*(5), 445–470. doi:10.1111/j.1532-7078.2009.00027.x

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77. doi:10.1016/0364-0213(90)90026-S

Markman, E. M. (1994). Constraints on word meaning in early language acquisition *. *Lingua*, *92*, 199–227. Retrieved from http://www.sciencedirect.com/science/article/pii/0024384194903425

Mather, E., & Plunkett, K. (2010). Novel labels support 10-month-olds' attention to novel objects. *Journal of Experimental Child Psychology*, *105*(3), 232–242. doi:10.1016/j.jecp.2009.11.004

Mather, E., & Plunkett, K. (2011). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language*, *38*, 933–950. doi:10.1017/S0305000910000401

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*(6), 1288–1301. doi:10.1111/j.1467-8624.1991.tb01606.x

Mervis, C. B., Golinkoff, R. M., & Bertrand, J. (1994). Two-year-olds readily learn multiple labels for the same basic-level category. *Child Development*, *65*(4), 1163–1177. doi:10.2307/1131312

Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *Journal of Child Language*, *22*(1), 19–46. doi:10.1017/S0305000900009612

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. doi:10.1177/0956797612438734

Rescorla, L. A. (1980). Overextension in early language development. *Journal of Child Language*, *7*(2), 321–335. Retrieved from http://journals.cambridge.org/action/displayAbstract?fromPage=online{\&}aid=1766684{\&}fulltextType=RA{\&}fileId=S0305000900002658

Shipley, E. F., Smith, C. S., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. *Language*, *45*(2.1), 322–342. doi:10.2307/411663

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. doi:10.1016/S0010-0277(00)00081-0

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. doi:10.1111/1467-9280.00485

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, *46*(1), 39–56. doi:10.1006/jmla.2001.2799

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175. doi:10.1111/1467-9280.00127

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. doi:10.1111/j.1532-7078.2011.00084.x

Weatherhead, D., & White, K. S. (2016). He says potato, she says potahto: Young infants track talker-specific accents. *Language Learning and Development*, *12*, 92–103. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/15475441.2015.1024835

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. doi:10.1016/j.jml.2008.03.001

Wojcik, E. H., & Saffran, J. R. (2013). The ontogeny of lexical networks: Toddlers encode the relationships among referents when learning novel words. *Psychological Science*, *24*(10), 1898–1905. doi:10.1177/0956797613478198
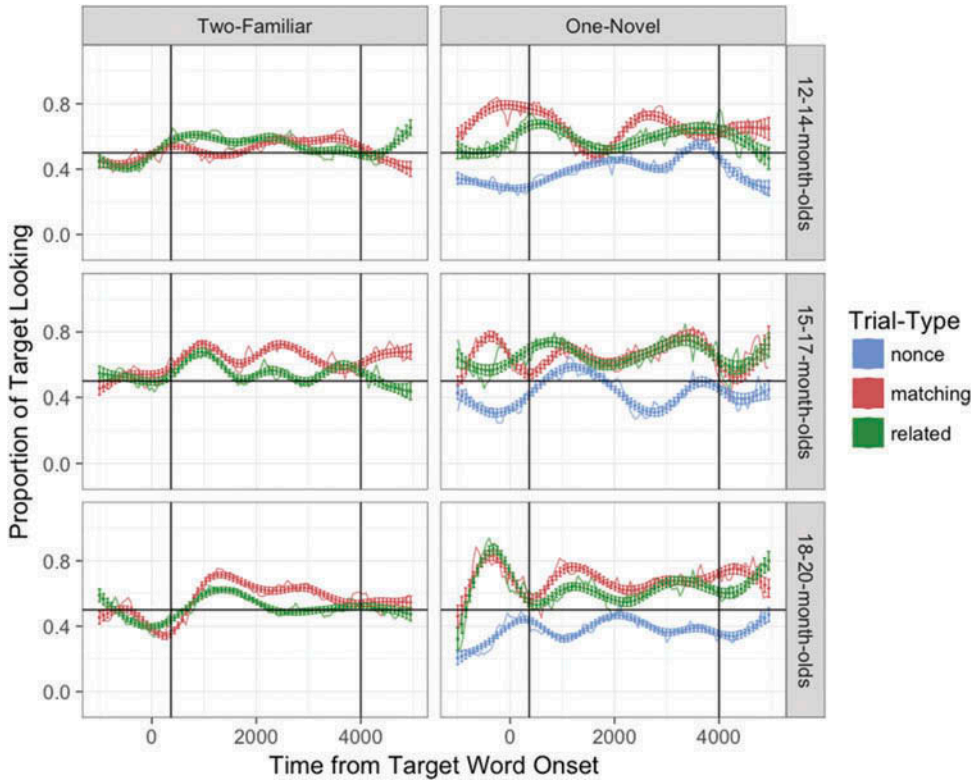
Appendix

Figure A1. Timecourse of infant gaze by trial-type and age-group. This figure shows the proportion of fixation to the target image over time, separately for each age-group (horizontal panels), pair-type (vertical panels) and trial-type (color). For instance, the bottom left cell shows 18–20-month-olds target looking on two-familiar pair-type trials (e.g., foot-juice), with the red line reflecting looking when they heard the matching audio target (e.g., foot), and the green line reflecting looking upon hearing the related audio target (e.g., sock). We smooth over the sample-level data (thin lines) from the eyetracker, averaging over subjects and trials, and adding 95% bootstrapped CIs (thicker line). The black vertical lines demark the target window of analysis from 367–4000 msec. The baseline window of analysis is all looking before target onset, i.e., time 0. This figure is provided for data transparency, analyses were collapsed over delineated windows of analysis; see text for further details.