HHS Public Access

Familiarity plays a small role in noun comprehension at 12–18 months

Hallie Garrison[1], **Gladys Baudet**[1], **Elise Breitfeld**[2], **Alexis Aberman**[3], **Elika Bergelson**[1]

[1]Duke University, Durham, North Carolina

[2]Princeton University, Princeton, New Jersey

[3]Cornell University, Ithaca, New York

Abstract

Infants amass thousands of hours of experience with *particular* items, each of which is representative of a broader category that often shares perceptual features. Robust word comprehension requires generalizing known labels to new category members. While young infants have been found to look at common nouns when they are named aloud, the role of *item familiarity* has not been well examined. This study compares 12- to 18-month-olds' word comprehension in the context of pairs of their own items (e.g., photographs of their own shoe and ball) versus new tokens from the same category (e.g., a new shoe and ball). Our results replicate previous work showing that noun comprehension improves rapidly over the second year, while also suggesting that item familiarity appears to play a far smaller role in comprehension in this age range. This in turn suggests that even before age 2, ready generalization beyond particular experiences is an intrinsic component of lexical development.

1 | INTRODUCTION

While young infants do not produce words until around the first birthday (Fenson et al., 1994), they begin to understand words months earlier, particularly for common nouns and salient individuals (Bergelson & Swingley, 2012; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999). One challenge in this process is that word learning is necessarily both specific and general. That is, infants must be able to not only link their experiences with *particular* referents in the world and the words used to label them, but also properly generalize their knowledge to *new* instances of the word. It is not enough for a learner to realize that their favorite toy is called "ball" when a helpful caretaker unambiguously labels it this way. First of all, the world rarely provides such clear-cut data. But even if it did, learners would still require some sense of the general concept or category to enable proper extension to new instances. For instance, if a baby can pick a ball out of a basket of toys they have never

encountered upon hearing "ball," or recognize a ball from a 2-dimensional drawing, this provides evidence that they know "ball" refers to a category of round rollable toys.

Indeed, behavioral laboratory studies of word comprehension capitalize on this logic, testing whether young infants look more at a named image when presented with multiple options, in what is often dubbed the "looking-while-listening" task (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). Given infants' success in this task from 6 months onward with both novel images and videos of familiar categories or people (Bergelson & Aslin, 2017a; Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012), they must already have some rudimentary skills to conceptualize and categorize word forms and referents. In the current work, we investigate how broad of a category infants have represented, and how their own experience with particular tokens comes into play.[1]

Relevantly, categorization skills at both a conceptual and perceptual level improve alongside growing linguistic skills over infants' first year of life (Oakes, n.d.). While infant categorization results are strongly influenced by methods and stimuli (Rakison & Yermolayeva, 2010), a growing literature shows that infants' processing of specific tokens can shape aspects of their discrimination, attention, memory, and learning of categories, within and across animate and inanimate classes such as people, monkeys, pets, and vehicles (Oakes, n.d.). For instance, by 6 months, infants treat previously unseen instances of a visual category that are the average of previously seen tokens as familiar (de Haan, Johnson, Maurer, & Perrett, 2001), and also remember individual instances (Oakes & Kovack-Lesh, 2013). Furthermore, both longer-term home experience and brief laboratory experience influence new category formation (Bornstein & Arterberry, 2010). With common categories like animals and furniture, previous research has shown that over 12–30 months, infants' categorization shifts from more to less inclusive, with a large role for perceptual contrast in explaining behavior (Bornstein & Arterberry, 2010). Finally, while many aspects of categorization are perceptually guided, by age 1 infants can group objects by unseen conceptual features like function as well (Träuble & Pauen, 2007). On some accounts, words are one such grouping feature (Deng & Sloutsky, 2015). Indeed, whether words are a privileged type of feature is under (at times vociferous) debate (Althaus & Westermann, 2016; Ferguson & Waxman, 2017).

Views on how infants come to properly extend words to new instances largely fall in two camps: narrow-to-broad and broad-to-narrow (Hirsh-Pasek, Golinkoff, Hennon, & McGuire, 2004). On the narrow-to-broad view (Hirsh-Pasek et al., 2004; Smith, 2000), children are believed to first learn that a word is used to label individual objects and then, over time, begin to generalize the word to a category of objects (on either perceptual or other bases). On this view, children construct categories inductively—they begin to group objects together based on similarities they notice over time, eventually having the second-order insight that Xs are X-shaped for many concrete nouns X (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Conversely, the broad-to-narrow view of word learning suggests that infants are biased to understand a noun as referring to a category of like objects, before later

---

[1] We focus on meanings, but *mutatis mutandis*, the same issues apply to early phonetic representations.

understanding more fine-tuned, specific references (Csibra & Shamsudheen, 2015; Waxman & Markow, 1995). However, much of this work has used novel words and objects, leaving a bit unclear how these processes unfold at a real-life scale, that is, during infants' daily learning experiences with common words and referents.

Muddying the waters further, language can of course label individuals or categories, often leaving a grammatical signature of this distinction. For instance, English syntax generally separates proper nouns (for which there is no broader category) and common nouns by adding articles for the latter ("here's Sarah" vs. "here's *a* ball"). However, early on, caretaker labels such as "mommy" and "daddy" are proper-noun-like in their extension ("Here's Mommy!"). Indeed, for 6-month-olds, these words refer to their mother and father and not other women and men (Tincoff & Jusczyk, 1999). Moreover, infants' early productive vocabulary contains both nouns naming individuals (e.g., "mommy") alongside count nouns such as "dog" (Fenson et al., 1994). Building on this, exciting recent evidence suggests infants know both individual- and category-level nouns from the onset of word comprehension around six months (Campbell, 2018), supporting a "narrow-*and*-broad" view.

The current investigation extends these lines of research to further explore the effect of infants' linguistic and physical environments on their noun and object knowledge, and how this may change across development. We ask whether infants understand words better when tested with images of the concrete nouns they see in their own daily lives, as opposed to other, laboratory-selected images of these same words. We look at this between 12 and 18 months, a particularly rich developmental period when the productive vocabulary is beginning to grow, comprehension rapidly improves, and infants' ability to discriminate instances of common categories is well established (Bergelson & Swingley, 2015; Fenson et al., 1994; Oakes, n.d.).

Using the looking-while-listening spoken word comprehension task (Fernald et al., 2008), we measure whether infants' performance varies as a function of whether the images they see are of their *own* objects from home, or *other* exemplars of the same category that are typical, but novel. While visual processing and categorization have been shown to vary as a function of home experience with objects or pets by 5–6 months (Bornstein & Arterberry, 2010; Hurley, Kovack-Lesh, & Oakes, 2010), this has not, to our knowledge, been tested over a set of common nouns during spoken word comprehension (cf. Campbell, 2018, which we return to in the discussion).

Given that previous word comprehension studies using this method with laboratory-selected images find that word comprehension is relatively robust at 12–18 months (Bergelson & Swingley, 2012, 2015; Swingley, Pinto, & Fernald, 1999), we predict infants will readily understand words when shown novel exemplars (i.e., in the "other-image" trial type), looking to the named image significantly more than the distracter. Whether they understand words equivalently or even better when shown images of their own objects is the open question we test here; we have no reason to think performance would be *worse* in the own-image trial type based on the previous literature. If understanding a word—as measured in this task—is tightly linked to infants' own experiences with referents in their daily life, we

would predict better performance on their own items over other ones. In contrast, if understanding a word entails recognizing any reasonable instance, whether new or familiar, we would predict equivalent and above-chance performance across both other-image and own-image trial types.

Equivalent performance on both trial types would be more consistent with either a broad-to-narrow view of word learning or a narrow-*and*-broad view (cf. Campbell, 2018), while better performance with their own images than the laboratory-selected other images would better align with the narrow-to-broad view, suggesting that infants' (initial) understanding of the referential nature of words is tied to their specific learning instances. A further possible interpretation of equivalent performance across trial types would be that infants do undergo a shift consistent with the narrow-to-broad view, but that this shift has concluded before age 1. Critically, performance across and/or within trial types may vary with age; we test a cross-sectional sample to address this. Given that word comprehension improves over infancy, with particular boosts in comprehension using the present method just after the first birthday (Bergelson & Swingley, 2012, 2013, 2015); here, we investigate whether this improvement may be due to better categorization and generalization skills interacting with linguistic development.

Thus, the present study asks two questions: (a) Do infants understand words better when tested with images of their own items versus other novel ones? And (b) does comprehension in the context of such own or other images change as a function of age over 12–18 months?

2 | METHODS

All audio and visual stimuli, data, and code are on OSF: https://osf.io/pb2g6/. We report all measures and manipulations collected, along with all trial or participant exclusion information below.

2.1 | Participants

Our final sample included thirty 12- to 20-month-olds (range: 11.56–18.23 months, $M =$ 14.78, 13 girls). An additional 5 infants were excluded due to fussiness, poor calibration, or parental interference resulting in usable data for <50% of trials (see trial exclusion criteria below, cf. Bergelson & Aslin, 2017b for prior use of this exclusion approach).[2] Sample size was determined based on prior language development research, which typically has 16 infants/study. Given our novel manipulation, we sought to double this sample size to 32, but after exclusions (as delineated below) were left with 30 infants (cf. Frank et al., 2017).

Infants were recruited from a metropolitan area in the southeast United States by mail, phone, email, and in person. This study was conducted in accordance with the guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by Duke University's IRB. All of the

---

[2]The researcher running the experiment noted infants' fussiness or need to end the study early (e.g., crying and squirming) in a log immediately after the study; they were not privy to any type of result summary, basing notes on infants' behavior and eyetracker track-loss signals. These notes were then cross-checked with our data-driven exclusion at the trial level, as detailed below.

participants were healthy, full-term, had no history of vision or hearing loss, and heard English at least 75% of the time. Based on demographic forms completed by families, the current sample primarily came from upper-middle class homes; 86.67% of mothers completed a bachelor's degree or higher, 43.33% of mothers worked 40 hr/week or more, 50% had a family income of>$100 k/year, and the majority were White.

2.2 | Materials

2.2.1 | Visual stimuli—For the *own-image* trial type, parents provided 12 photographs of items their child commonly sees, based on a suggested list of 26 items. Suggested items included toys, clothing, pets, and food-related objects. Corresponding *other* images were selected from a laboratory database that is part of the SEEDLingS Corpus (Bergelson, 2016), after ensuring that the laboratory's image depicted a different object than parents provided. For instance, if the "book" picture from the family and the laboratory database depicted the same book, a different book image was used. Four warm-up trial images were also displayed, including apple, bottle, spoon, bear, and cat; warm-up items were chosen so that they did not overlap with experimental stimuli. All images were edited onto a plain gray background (see Figure 1).

2.2.2 | Item pairing—We tailored test items individually. Specifically, of the twelve images parents provided, eight items (corresponding to eight target words) were combined into four yoked item pairs for each infant. Items were chosen based on photograph clarity and two pairing criteria: (a) avoid pairing items with the same initial sound, and (b) avoid creating mixed animacy pairs (e.g., duck/pacifier) given that infants are highly drawn to faces. In two cases, it was not possible to avoid overlapping onset (bottle/blanket and block/ball; 1.67% of item-pairs). In six cases (5% of item pairs), it was necessary to pair animates with inanimates.

2.2.3 | Audio stimuli—Sentences containing each target word were recorded by a female using infant-directed speech. Each sentence was 1.1–1.8 s long and normalized to 72 db using Praat. The sentences were in one of four constructions: "Can you find the X?"; "Do you see the X?"; "Look at the X!"; and "Where is the X?" (X being the target object word). One carrier phrase was used per item pair. During warm-up trials, sentences played over computer speakers. During test trials, only the parent heard these pre-recorded sentences; infants heard them from their parent (as detailed in Procedure).

2.2.4 | Questionnaires—Parents completed the MacArthur-Bates Communicative Development Inventory (CDI) Words & Gestures Form (Fenson et al., 1993) and a Word Exposure Survey, which asked parents to report how often they believed their child heard the eight words tested in the experiment on a 5-point scale (1 = "never" and 5 = "several times a day"). Parents were also asked to describe infants' vocalizations and to give examples of words their child could say, if their child had begun to talk. Researchers selected "Saying Words" if parents named five or more words that their child could produce. Finally, parents were asked whether their child was breast-fed, bottle-fed, both, or neither, in order to establish whether "milk" would be an appropriate item to display in an image.

2.3 | Procedure

Infants' word comprehension was tested using an adapted looking-while-listening procedure (Bergelson & Aslin, 2017a; Bergelson & Swingley, 2012). After paperwork was complete, infants and caretakers were escorted to a dimly lit testing room, where infants were seated on their parent's lap in front of a computer monitor (33.7×26.9 cm screen, $1,280 \times 1,024$ resolution). An Eyelink 1,000+ Eyetracker camera (SR Research), mounted on a moveable arm and set to "remote" mode, collected infants' visual fixation data, sampling monocularly at 500 Hz ($<50°$ average accuracy), aided by a small, high-contrast sticker on infants' forehead. After calibration, infants were shown four warm-up trials, during which a single image appeared on the screen, while a pre-recorded sentence directed their attention to the object.

After this warm-up, infants would see 32 test trials, half in each trial type (own and other). In each test trial, infants saw photographs of two objects on a gray background, one on each side of the screen, displayed at 500×500 pixels. Each of the 8 target items was tested on four separate trials, two per trial type (e.g., if ball was a test item, it was the target twice in an own-image trial and twice in an other-image trial). Parents, who wore headphones and a visor to block their view of the screen, heard test sentences over headphones, which they repeated aloud to the infant when prompted by a beep. The target word occurred 3–3.5 s after the trial began. An experimenter pressed a button to indicate when the parent said the target word, after which the images were shown for 5 s (see Figure 2).

Trial order was pseudorandomized such that neither the same item pair nor the same carrier phrase appeared back-to-back, and target images did not occur more than twice in a row on the same side of the screen. Additionally, whether an image appeared on the left or right was controlled within item pairs across trials. As needed (i.e., when infants appeared fussy or distracted), a 2-s video including colorful shapes, smiley faces, and a bird-chirping sound was invoked between trials to maintain infant interest.

2.4 | Data analysis

We used R (version 3.6.1; R Core Team, 2019),[3] for all our analyses; all scripts and data used to generate this paper are currently on OSF (see link at start of Methods section).

3 | RESULTS

We present three sets of analyses below, after providing descriptive statistics from the questionnaire data; the first two analyses were planned, while the third was exploratory. First, to look at whether infants word comprehension is influenced by own versus other images, and whether this varies as a function of age, we examine the proportion of target

---

[3]We, furthermore, used the R packages *AICcmodavg* (version 2.2.2; Mazerolle, 2019), *broom* (version 0.5.2.9002; Robinson & Hayes, 2019), *citr* (version 0.3.2; Aust, 2019), *dplyr* (version 0.8.3; Wickham et al., 2019), *forcats* (version 0.4.0; Wickham, 2019a), *ggplot2* (version 3.2.1; Wickham, 2016), *kableExtra* (version 1.1.0; Zhu, 2019), *knitr* (version 1.26; Xie, 2015), *lme4* (version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015), *lubridate* (version 1.7.4; Grolemund & Wickham, 2011), *Matrix* (version 1.2.18; Bates & Maechler, 2019), *papaja* (version 0.1.0.9842; Aust & Barth, 2018), *polypoly* (version 0.0.2; Mahr, 2017), *purrr* (version 0.3.3; Henry & Wickham, 2019), *readr* (version 1.3.1; Wickham, Hester, & Francois, 2018), *readxl* (version 1.3.1; Wickham & Bryan, 2019), *stargazer* (version 5.2.2; Hlavac, 2018), *stringr* (version 1.4.0; Wickham, 2019b), *tibble* (version 2.1.3; Müller & Wickham, 2019), *tidyr* (version 1.0.0; Wickham & Henry, 2019), *tidyverse* (version 1.3.0; Wickham, 2017), and *TOSTER* (version 0.3.4; Lakens, 2017).

looking, that is, our proxy of comprehension, considering the influence of trial type and age, within a window of interest. Next, in order to determine whether the real-time process of word comprehension unfolds differently across own versus other images, and across age, we compare the target-looking trajectories directly using the growth curve analyses. Finally, to explore the possibility that our results are influenced by the similarity between images provided by infants' parents and laboratory-selected images, we incorporate similarity ratings from adults into our eyetracking analysis.

3.1 | Questionnaire results

For the CDI, parents reported that infants understood 157.60 words on average ($SD = 89.96$, range = 25–335) and produced 25.20 words ($SD = 35.64$, range = 0–137). About 66.67% of infants were reported to have said at least five words. For the Word Exposure Survey, where we asked how often parents believed their child heard each of the target words, parents reported that these words were quite common: 82.08% of responses were "Once a Day" or "Several Times a Day," that is, a 4 or 5 on a 5-point scale (see Table 1). For the query about word production, convergent with the CDI production results, 70% of infants were reported to have started saying words. About 53.33% of infants were bottle-fed or bottle- and breast-fed.[4]

3.2 | Eyetracking data preparation and exclusion

After exporting the eyetracking data from Eyelink, we converted the fixation data into 20-ms bins, aligned such that time 0 reflected the time of target word onset (e.g., "ball" in "Can you find the ball?"). For each trial, we then computed the proportion of target looking over a 367- to 2,000-ms target window of analysis, that is, from the first eye movements that may be relevant to the target word (Fernald et al., 1998) to 2 s later (as in Swingley & Aslin, 2002).[5]

Our final dataset included 831 trials from 30 infants ($M = 27.70/32$ trials, $SD = 4.36$, range:16–32). To arrive at this final dataset, we excluded trials in which parents failed to say the target word ($n = 4$), or in which infants failed to look at either the target or distracter for more than 1/3 of the 367- to 2,000-ms window of interest ($n = 210$ low-data trials). This latter type of trial exclusion generally occurred if infants failed to look at the screen during a given trial, and/or because the experiment was ended early due to infant fussiness/crying. The experiment ended early for four infants; no experiment ended before reaching 27 of 32 trials. Following this trial-level exclusion, we then excluded participants who contributed data to <50% of trials ($n = 5$); these participants accounted for 98 of the 210 low-data trials (cf. Bergelson & Aslin, 2017b).

---

[4]In practice, parents sometimes indicated 'breast-fed only' when infants received breastmilk by bottle. We thus retained the item 'milk' regardless of feeding status in the case ($n = 1$) that parents sent us a photograph of a bottle of milk as one of the child's items.
[5]While this proportion is sometimes subtracted from a pre-target baseline to account for saliency differences among images, here such a correction was unnecessary as the proportion of target looking in the pre-target baseline from –2,000 to 0 ms did not differ from chance (0.5) by Wilcoxon test (pseudomedian = 0.50, $p = .38$; a Wilcoxon test was used because a Shapiro test and Q–Q plot indicated that the distribution was highly divergent from a normal one, $p < .001$).

3.3 | Influence of age and image familiarity on overall comprehension

Testing whether age and referent familiarity influenced word comprehension, we first conducted a multi-level linear regression testing whether (centered) age and trial type (own image/other image) accounted for significant variance in the mean proportion of target looking, with a random effect for subject (i.e., mean proportion of target-looking age-centered + trial type + (1|subject)); model residuals were normal by a Shapiro test ($p = .82$). We found an age effect, but no effect of trial type. Comparing with models without trial type or with an age*trial-type interaction, we found that trial type did not significantly interact with age ($p = .51$) or improve fit above a model with age alone ($p = .42$; see Table 2 for model summaries).

Given the significant effect of age in our data, we were curious to explore whether trial-type differences might emerge in younger versus older infants. We thus conducted a median-age split (14.69 months), which created two age-groups ($n = 15$/group). All patterns reported above with continuous age hold for age-group (i.e., the older half showed stronger comprehension than the younger half, with no effects of trial type; see Figures 3 and 4).

Performance on other versus own trials did not vary overall by a paired two-tailed t test (mean difference = -0.02, $t(29) = -0.80$, $p = .43$), or within either age-group (*older*: mean difference = 0.003, $t(14) = 0.12$, $p = .90$); *younger*: mean difference = -0.04, $t(14) = -1.09$, $p = .30$).[6]

In order to more thoroughly dissect the results above, we conducted equivalence testing in addition to the paired and one-sample t tests (the latter of which only let us reject or fail to reject the null, rather than establish whether we can reject or fail to reject an effect greater than the smallest effect of interest (Lakens, McLatchie, Isager, Scheel, & Dienes, 2020)). Specifically, this lets us investigate (a) whether the null result for other-image trials in the younger infants was statistically equivalent to zero and (2) whether the lack of trial-type difference we found overall and in each age-half was statistically equivalent to zero. In all cases, we determined the equivalence bounds based on the effect size we would be able to detect at $\alpha = 0.05$, power = 0.8, with our sample size (i.e., 15 for each age-half or 30 overall).

For the comparison to chance for the younger group, the equivalence test was not significant ($t(14) = -1.40$, $p = .092$, equivalence bounds: ±0.084). Combined with the nonsignificant null hypothesis test above, this means that the results for the younger group on the other-image trial type are inconclusive; we return to this point in the discussion.

For the trial-type difference, in all three cases (overall and for each age-half), the equivalence test was significant (*overall*: $t(29) = 2.10$, $p = .022$, equivalence bounds: ±0.066; *older*: $t(14) = 3.07$, $p = .007$, equivalence bounds: ±0.079; *younger*: $t(14) = 1.86$, $p = .042$, equivalence bounds: ±0.108). Combined with the nonsignificant null hypothesis test reported above for each of these cases, this lets us conclude that the observed effect *both*

---

[6]The distribution of mean proportion of target looking differed slightly from a normal distribution ($p = .04$); for simplicity, we report paired t tests rather than Wilcoxon tests but all patterns hold for paired Wilcoxon tests.

does not differ from zero *and* is statistically equivalent to zero: performance on own- and other-image trials in this study was the same, as measured by proportion of target looking.

Taken together, our age-half analysis is consistent with the model results in finding no statistical support for a trial-type difference when examining the proportion of target looking over a set window (though we note that younger infants in the other-image trial type showed slightly weaker performance by *t* test; see Table 3). Binomial tests rendered a consistent pattern: Most infants' means were above chance in both age-groups in the own-item trial type, as were older infants' in the other-item trial type (see Table 3 for Ns, means, *t* test, and binomial test results).

3.4 | Growth curve analysis

While the subject-level aggregated data showed clear age effects but no trial-type effects, the time course data (Figure 4) highlighted diverging gaze patterns across trial type over time, especially in the younger infants. To further investigate the online dynamics of spoken word comprehension in this study, we used growth curve analysis over our 367- to 2,000-ms analysis window. Following previous research, we transformed the data from its raw categorical, bounded state (i.e., a 1 or 0 for target looking vs. not in any given timebin) to continuous unbounded values using the empirical logit of target looking per subject per trial type per timebin (cf. Mahr, McMillan, Saffran, Weismer, & Edwards, 2015). This in turn made the standard linear statistics available to us (Barr, 2008).

Given issues with model convergence and singularity, and the unambiguous role of age in our results, we ran separate growth curve models for younger and older infants. This allowed us to focus in on the effect of trial type in predicting (empirical logit-transformed) proportion of target looking across the time terms. We followed the approach outlined by Mirman (2014) and Mahr et al. (2015), and used four orthogonal polynomial time terms, examining their interaction with trial type.[7] We also included random slopes and intercepts for each subject, and the interaction of trial type with each time term (with other-image trial type as the reference condition). Statistical significance for individual parameter estimates was assessed using the normal approximation (cf. Mirman, 2014). As seen in Figure 5, the models provided a reasonably good fit to the data.

For the younger infants, the key results of this analysis were a main effect of trial type, an interaction of trial type with the linear, quadratic, and quartic time terms. That is, younger infants performed better in the own-image condition, and looking patterns showed divergent average curve height, slope, and inflectional steepness across conditions. More descriptively, infants appear to switch their gaze between target and distracter more for laboratory-selected other-image trials, leading to a more "wiggly" looking-time curve. For the older infants, consistent with the previous analysis collapsing time, the growth curve analysis revealed no main effect of trial type, though there was a significant trial-type interactions with the linear and cubic time terms, that is, older infants' looking showed slope and curve shape differences across trial types. More descriptively, older infants got to the target on own-image trials a little faster, but spent a bit less time there. Thus, considering how eye

---

[7]cf. this very useful vignette http://www.eyetracking-r.com/vignettes/growth_curve_analysis.

movements unfolded over time in this growth curve analysis revealed a role for trial type, particularly within the younger infants in this sample; the role of trial type remained modest (i.e., there was no main effect of trial type) for the older infants (see Figure 5 and Tables 4 and 5 for model fits and summaries).

3.5 | Similarity ratings

One possible contributor to infants' comprehension is the similarity between the parent-provided images and laboratory-selected images they were tested on. That is, although in a given trial both images that infants saw were either their own items or other laboratory-selected images, it is possible that for some infants, the laboratory-selected items were more similar to their own items than for others. To directly test this, in an exploratory follow-up analysis, we had adult volunteers rate all images that the full sample of infants was tested on. Specifically, adult volunteers were shown pairs consisting of a parent-provided image and a laboratory-selected image of the same object (e.g., one infant's own ball and the other ball image they were tested on) displayed side by side on a gray background. For each pair, volunteers were asked "How similar are these two items," and prompted to select a number on a 7-point Likert scale (1 = similar to 7 = different). Researchers solicited participation in this online qualtrics survey through emails to individuals who did not work in the laboratory and who did not know any details about the project (see Table 6).

After removing false starts from the qualtrics data, 20 adults completed the rating task, providing ratings for 5,912 pairs (out of 5,920; there were eight missing ratings). We then removed ratings for items from the five infants who were excluded due to contributing insufficient data above, along with two further infants (due to an experimenter error, adults also rated items for two infants for whom an experiment was prepared but who never came in). In the final dataset of 5,912 rated pairs, the grand mean across similarity ratings was 3.53 ($SD = 1.95$). Individuals raters' scores were normally distributed around this mean (range = 1.84 to 5.47; $p = .69$ by Shapiro test). Ratings on individuals items (e.g., infant subject 1s own vs. other "ball" images) were more broadly and not normally distributed ($R = 1.40$–6.47, $p < .001$ by Shapiro test).

We then incorporated these ratings into our eyetracking data analysis, calculating a similarity score for each infant. This score was the average similarity score from our adult volunteers for each infants' eight own and eight other items. For the window-of-interest analysis, there was no correlation between average similarity ratings and infants' overall subject means ($R = 0.10$, $p = .58$), or between similarity ratings and their subject means on either trial type (other: $R = 0.13$, $p = .48$; own: $R = 0.03$, $p = .87$). Moreover, a model adding similarity score as a predictor (above and beyond age) did not fit the data better than one that omitted this term ($p = .78$).[8] Thus, this final analysis suggests that the visual similarity between laboratory-selected and parent-provided items did not drive infants' performance in our eyetracking study, though given its exploratory nature, replication is called for before strong conclusions can be made on this point.

---

[8]Unfortunately, issues with model singularity (likely due to overfitting and data sparsity) prevented us from incorporating similarity scores into our growth curve analyses.

4 | DISCUSSION

We set out to answer two questions: Whether referent familiarity contributes to word comprehension, and how this may shift over the first half of the second year. Summarily, while we clearly see stronger comprehension in older infants across year 2 (consistent with the literature, e.g., Fernald et al., 1998), we found only a modest role for referent familiarity (particularly in our multi-level linear regression and paired trial-type comparisons). That is, we found evidence for word comprehension across our sample when infants saw images of referents from their daily lives, and novel images of those same kinds of items. Convergently, we found statistically equivalent performance across these two types of images (i.e., own vs. other trials) in our proportion-based analyses, alongside some finer-grained differences in the online dynamics of word comprehension.

There are several possible interpretations of this pattern of results. One possibility is that infants could not actually tell the familiar and novel images of our referents apart. Under this scenario, infants' statistically equivalent performance on own versus other trials in our proportion-based analysis would suggest not that infants generalize to new tokens, but that they have relatively poor representations of referents they have hundreds of hours of experience with. We find this unlikely for several reasons.

First, our growth curve analyses (which probed the real-time dynamics of word comprehension) found that younger infants' fine-grained looking patterns varied between trial types; it seems unlikely that a difference apparent to younger infants would be unapparent to older ones. Second, in the object categorization domain, even 6-month-olds can detect subtle changes in a two-shape visual array, for example, when shown a diamond and a circle on a gray background, they notice if the diamond changes color after a brief blank screen interlude (Cantrell, Kanjlia, Harrison, Luck, & Oakes, 2019). These are far subtler differences than those between our parent-provided and laboratory-selected versions of items, and our sample is 6–12 months older, suggesting it is quite unlikely they would not be able to tell an image of their dog or pacifier apart from a different token of these categories (cf. Feigenson, 2005). Relatedly, six-month-olds also readily recognize their own mother and father as compared to other women and men in a word comprehension context (Campbell, 2018; Tincoff & Jusczyk, 1999). Thus, previous work provides support for the idea that using similar methods, infants can tell when referents change, and "their" version of a referent apart from a different one.

Third, even more compelling evidence that the infants in our study could tell their items and the new items apart comes from a recent study on pets, proper nouns, and count nouns (Campbell, 2018). In their within-subjects study, 12- to 15-month-olds saw images of either their pet and another same-breed pet (e.g., their poodle Joey and a novel poodle), or their pet and a different-species pet (e.g., Joey and a novel cat) and heard "dog," "cat," or "Joey" on different trials. Infants looked at the correct referent across the board: On hearing, for example, "Joey," they looked at their own pet whether it was next to another same-breed dog, or a cat; when hearing "cat" or "dog," they looked at the cat or dog when only one of each was on the screen. When hearing "dog" while seeing their own pet and a novel same-breed dog, they looked at both equally. Thus, Campbell's findings provide evidence that

infants in the age range tested here can tell their own items apart from new ones (since, e.g., "Joey" only applying to the former) and, moreover, understand that count nouns such as "dog" apply to both. More generally, this supports a *narrow-and-broad* account of word learning, where both individuation and categorical knowledge are part of early word learning.

If we had seen that infants only considered their own familiar items as appropriate referents, that would have provided evidence of overly narrow categories. This is not what we find. If anything, previous studies suggest that rather than being overly specific in their representations, in some contexts infants are overly accepting, that is, include clearly inappropriate referents as instances of a word. Indeed, previous work suggests that infants' representations for concrete nouns go through fine-tuning and competition depending on perceptual and conceptual features (Arias-Trejo & Plunkett, 2010; Bergelson & Aslin, 2017a, 2017b). For instance, one study finds that around 12 months, infants look at, for example, an image of "juice" when hearing juice or "milk"; by around 18 months, they are less likely to do so (Bergelson & Aslin, 2017b). This suggests that in some circumstances, even when a given image is definitely not a category member (i.e., far more than simply being an unfamiliar but plausible token), early in year 2, infants treat it as a plausible referent. In this light, it is perhaps not surprising that in the present study, images of infants' own common noun referents and images of sample referents from the same category were both considered plausible referents to infants.

Taken together, these previous studies leave it unlikely that infants simply could not tell familiar and novel instances of referents apart here. To us, a more parsimonious interpretation is that own-referent familiarity is both detectable and largely irrelevant to this spoken word comprehension task. In this way, referent familiarity appears to be analogous to voice familiarity. For instance, while in certain contexts young infants can discriminate novel versus familiar voices (Johnson, Westrek, Nazzi, & Cutler, 2011), they do not show a decrement in spoken word comprehension for familiar nouns when hearing novel talkers (Bergelson & Swingley, 2018), ostensibly because talker identity is not relevant to the task. Here too, with common nouns and typical images of referents, spoken word comprehension appears to proceed regardless of image familiarity.

To be clear, we are not claiming that referent familiarity could *never* play a large role in spoken word comprehension. We are suggesting that when using typical images of common nouns, in a simple spoken word comprehension task with 12- to 18-month-olds, it appears to play a relatively minor role, especially for infants approaching 18 months. Just as listeners *can* take talker identity into account (e.g., showing surprise at hearing a young child discussing cigarette smoking; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008), so too could familiarity play a big role in the right circumstances. For instance, typicality appears to drive comprehension in 12-month-olds to a greater degree than in older infants (Meints, Plunkett, & Harris, 1999). Given that typicality effects emerge from accrued experiences, for categories where infants have less experience, familiarity may play a more overt role. This awaits further work.

4.1 | What about younger infants?

Our reasons for testing 12- to 18-month-olds rather than younger infants were both practical and theoretical. Older infants know more words and know the words they do know better than younger infants by various metrics (Bergelson & Swingley, 2012, 2015; Fernald et al., 1998; Frank, Braginsky, Marchman, & Yurovsky, n.d.). Thus, in order to (a) be able to select a range of items for each child that we could pair while minimizing animacy mismatch or onset overlap, (b) have reasonable confidence that infants knew the labels for based on existing norms, results, and parental report, and (c) position ourselves better to see trial-type differences without worrying about potential floor or ceiling effects, that is, to capture a range of development where knowledge is strong but improvements are rapid, we opted for the first half of year 2.

However, our chosen age range leaves open the possibility that infants under 12 months would show a strong effect of item familiarity. Indeed, in our proportion-based analysis, younger infants in our sample showed inconclusive evidence of word comprehension in the other-item trial type, though their performance was statistically equivalent for own-item and other-item trials overall (as was older infants'). And although we found evidence of item familiarity (i.e., trial type) contributing differentially to the dynamics of real-time comprehension in our growth curve analyses (particularly for younger infants), we did not see such an effect in our multi-level analyses collapsing across our window of interest, in either age-group (or with age as a continuous variable). Our interpretation is that 12- to 15-month-olds' inconclusive performance on other-item trials is driven by lower levels of performance in this age-group in this task overall.

That said, we cannot rule out the possibility that younger infants than those tested here would show stronger performance on own versus other trials. However, here too we invoke data from Campbell (2018). That work finds that similar to the pet study described above, 6- and 9-month-olds do not look more at images of their own hand or ball when seeing them alongside novel ones and hearing "ball" or "hand," but they do look more at the their own hand or ball when hearing those words and seeing a second item from a different category (e.g., foot). We thus believe that we would fail to see large familiarity effects in younger infants, though this is an open empirical question.

5 | CONCLUSION

Taken together, these results support a growing literature showing that spoken word comprehension improves substantially over infancy, as infants quickly shift their attention to new appropriate referents in mere seconds as utterances unfold. Intriguingly, this work provides some evidence that referent familiarity appears to play a small role in this process, particularly for infants approaching the midpoint of year 2. In turn, this suggests that part of understanding words is robustly generalizing *beyond* thousands of hours of direct experience with specific referents, linking up categories and concepts with the words a given language uses to label them.

ACKNOWLEDGMENTS

## REFERENCES

Althaus N, & Westermann G (2016). Labels constructively shape object categories in 10-month-old infants. Journal of Experimental Child Psychology, 151, 5–17. 10.1016/j.jecp.2015.11.013 [PubMed: 26778468]

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. [PubMed: 19931091]

Aust F (2019). Citr: 'RStudio' add-in to insert markdown citations. Retrieved from https://CRAN.R-project.org/package=citr

Aust F, & Barth M (2018). papaja: Create APA manuscripts with R markdown. Retrieved from https://github.com/crsh/papaja

Barr DJ (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. Journal of Memory and Language, 59(4), 457–474.

Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. 10.18637/jss.v067.i01

Bates D, & Maechler M (2019). Matrix: Sparse and dense matrix classes and methods. Retrieved from https://CRAN.R-project.org/package=Matrix

Bergelson E (2016). Bergelson seedlings homebank corpus. 10.21415/T5PK6D.

Bergelson E, & Aslin R (2017a). Nature and origins of the lexicon in 6-mo-olds. Proceedings of the National Academy of Sciences of the United States of America, 114(49), 12916–12921. 10.1073/pnas.1712966114 [PubMed: 29158399]

Bergelson E, & Aslin R (2017b). Semantic specificity in one-year-olds' word comprehension. Language Learning and Development, 13(4), 481–501. [PubMed: 29200981]

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. 10.1073/pnas.1113380109 [PubMed: 22331874]

Bergelson E, & Swingley D (2013). The acquisition of abstract words by young infants. Cognition, 127(3), 391–397. 10.1016/J.COGNITION.2013.02.011 [PubMed: 23542412]

Bergelson E, & Swingley D (2015). Early word comprehension in infants: Replication and extension. Language Learning and Development, 11(4), 369–380. 10.1080/15475441.2014.979387 [PubMed: 26664329]

Bergelson E, & Swingley D (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. Child Development, 89(5), 1567–1576. [PubMed: 28639708]

Bornstein MH, & Arterberry ME (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. Developmental Psychology, 46(2), 350–365. 10.1037/a0018411 [PubMed: 20210495]

Campbell JCS (2018). The nature of infants' early object word comprehension (PhD thesis). University of British Columbia 10.14288/1.0371259

Cantrell LM, Kanjlia S, Harrison M, Luck SJ, & Oakes LM (2019). Cues to individuation facilitate 6-month-old infants' visual short-term memory. Developmental Psychology, 55(5), 905–919. [PubMed: 30702312]

Csibra G, & Shamsudheen R (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. Annual Review of Psychology, 66(1), 689–710. 10.1146/annurev-psych-010814-015232

de Haan M, Johnson MH, Maurer D, & Perrett DI (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. Cognitive Development, 16(2), 659–678. 10.1016/S0885-2014(01)00051-X

Deng W, & Sloutsky VM (2015). The development of categorization: Effects of classification and inference training on category representation. Developmental Psychology, 51(3), 392–405. 10.1037/a0038749 [PubMed: 25602938]

Feigenson L (2005). A double-dissociation in infants' representations of object arrays. Cognition, 95(3), B37–B48. [PubMed: 15788156]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, … Stiles J (1994). Variability in early communicative development. Monographs of the Society for Research in Child Development, 59(5), i–185. 10.2307/1166093 [PubMed: 8047076]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, & Reilly JS (1993). MacArthur-Bates communicative development inventories: User's guide and technical manual. San Diego, CA: Singular Publishing Group.

Ferguson B, & Waxman SR (2017). Linking language and categorization in infancy. Journal of Child Language, 44(03), 527–552. 10.1017/S0305000916000568 [PubMed: 27830633]

Fernald A, Pinto JP, Swingley D, Weinberg A, & McRoberts GW (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. Psychological Science, 9(3), 228–231. 10.1111/1467-9280.00044

Fernald A, Zangl R, Portillo AL, & Marchman VA (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children In Sekerina E, Fernández M, & Clahsen H (Eds.), Language acquisition and language disorders: Vol. 44. Developmental psycholinguistics: On-line methods in children's language processing (pp. 97–135). Amsterdam, The Netherlands: John Benjamins Publishing Company 10.1075/lald.44.06fer

Frank MC, Bergelson E, Bergmann C, Cristia A, Floccia C, Gervain J, … Yurovsky D (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. Infancy, 22(4), 421–435. [PubMed: 31772509]

Frank M, Braginsky M, Marchman V, & Yurovsky D (n.d.) Consistency in early language learning: The wordbank project. Cambridge, MA: MIT Press.

Grolemund G, & Wickham H (2011). Dates and times made easy with lubridate. Journal of Statistical Software, 40(3), 1–25. Retrieved from http://www.jstatsoft.org/v40/i03/

Henry L, & Wickham H (2019). Purrr: Functional programming tools. Retrieved from https://CRAN.R-project.org/package=purrr

Hirsh-Pasek K, Golinkoff RM, Hennon EA, & McGuire MJ (2004). Hybrid theories at the frontier of developmental psychology: The emergentist coalition model of word learning as a case in point. Weaving a Lexicon, (6 2014), 1–42. Retrieved from http://books.google.com/books?hl=en%7B&%7Dlr=%7B&%7Did=2Z-brX0N57lUC%7B&%7Doi=fnd%7B&%7Dpg=PR9%7B&%7Ddq=Weaving+a+lexicon%7B&%7Dots=R4SOoshhdq%7B&%7Dsig=QQx0VSI521Gfe24ISfT3Esbbk6I%7B%%7D5Cnhttp://books.google.com/books?hl=en%7B&%7Dlr=%7B&%7Did=2ZbrX0N57lUC%7B&%7Doi=fnd%7B&%7Dpg=PA173%7B&%7Ddq=Hybrid+Theories+at+the+Frontier+of+Devehttp://books.google.com/books?hl=en{\&}lr={\&}id=2ZbrX0N57lUC{\&}oi=fnd{\&}pg=PR9{\&}dq=Weaving+a+lexicon{\&}ots=R4SOoshhdq{\&}sig=QQx0VSI521Gfe24ISfT3Esbbk6I{\%}5Cnhttp://books.google.com/books?hl=en{\&}lr={\&}id=2ZbrX0N57lUC{\&}oi=fnd{\&}pg=PA173{\&}dq=Hybrid+Theories+at+the+Frontier+of+Deve

Hlavac M (2018). Stargazer: Well-formatted regression and summary statistics tables. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI) Retrieved from https://CRAN.R-project.org/package=stargazer

Hurley KB, Kovack-Lesh KA, & Oakes LM (2010). The influence of pets on infants' processing of cat and dog images. Infant Behavior and Development, 33(4), 619–628. 10.1016/J.INFBEH.2010.07.015 [PubMed: 20728223]

Johnson EK, Westrek E, Nazzi T, & Cutler A (2011). Infant ability to tell voices apart rests on language experience. Developmental Science, 14(5), 1002–1011. [PubMed: 21884316]

Lakens D (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. Social Psychological and Personality Science, 1, 1–8. 10.1177/1948550617697177

Lakens D, McLatchie N, Isager PM, Scheel AM, & Dienes Z (2020). Improving inferences about null effects with Bayes factors and equivalence tests. The Journals of Gerontology: Series B, 75(1), 45–57. 10.1093/geronb/gby065

Mahr T (2017). Polypoly: Helper functions for orthogonal polynomials. Retrieved from https://CRAN.R-project.org/package=polypoly

Mahr T, McMillan BT, Saffran JR, Weismer SE, & Edwards J (2015). Anticipatory coarticulation facilitates word recognition in toddlers. Cognition, 142, 345–350. [PubMed: 26072992]

Mazerolle MJ (2019). AICcmodavg: Model selection and multimodel inference based on (q)AIC(c). Retrieved from https://cran.r-project.org/package=AICcmodavg

Meints K, Plunkett K, & Harris PL (1999). When does and ostrich become a bird? The role of typicality in early word comprehension. Developmental Psychology, 35(4), 1072–1078. [PubMed: 10442875]

Mirman D (2014). Growth curve analysis and visualization using r. Boca Raton, FL: Chapman; Hall/CRC.

Müller K, & Wickham H (2019). Tibble: Simple data frames. Retrieved from https://CRAN.R-project.org/package=tibble

Oakes L (n.d.) Infant categories In Lockman J, & Tamis-LeMonda C (Eds.), The Cambridge handbook of infant development. Cambridge, UK: Cambridge University Press.

Oakes LM, & Kovack-Lesh KA (2013). Infants' visual recognition memory for a series of categorically related items. Journal of Cognition and Development, 14(1), 63–86. 10.1080/15248372.2011.645971

Parise E, & Csibra G (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. Psychological Science, 23(7), 728–733. 10.1177/0956797612438734 [PubMed: 22692337]

R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

Rakison DH, & Yermolayeva Y (2010). Infant categorization. *Wiley Interdisciplinary Reviews*: Cognitive Science, 1(6), 894–905. 10.1002/wcs.81 [PubMed: 26271785]

Robinson D, & Hayes A (2019). Broom: Convert statistical objects into tidy tibbles.

Smith LB (2000). Learning how to learn words: An associative crane In Becoming a word learner. New York, NY: Oxford University Press 10.1093/acprof:oso/9780195130324.003.003

Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, & Samuelson L (2002). Object name learning provides on-the-job training for attention. Psychological Science, 13(1), 13–19. Retrieved from https://pdfs.semanticscholar.org/22d3/cfbe61ca1ee306e99d0e0abef45f720a6a17.pdf [PubMed: 11892773]

Swingley D, & Aslin RN (2000). Spoken word recognition and lexical representation in very young children. Cognition, 76(2), 147–166. 10.1016/S0010-0277(00)00081-0. [PubMed: 10856741]

Swingley D, & Aslin RN (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science, 13(5), 480–484. [PubMed: 12219818]

Swingley D, Pinto JP, & Fernald A (1999). Continuous processing in word recognition at 24 months. Cognition, 71(2), 73–108. 10.1016/S0010-0277(99)00021-9 [PubMed: 10444905]

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

Tincoff R, & Jusczyk PW (2012). Six-month-olds comprehend words that refer to parts of the body. Infancy, 17(4), 432–444. 10.1111/j.1532-7078.2011.00084.x [PubMed: 32693484]

Träuble B, & Pauen S (2007). The role of functional information for infant categorization. Cognition, 105(2), 362–379. 10.1016/J.COGNITION.2006.10.003 [PubMed: 17129581]

Van Berkum JJ, Van den Brink D, Tesink CM, Kos M, & Hagoort P (2008). The neural integration of speaker and message. Journal of Cognitive Neuroscience, 20(4), 580–591. [PubMed: 18052777]

Waxman SR, & Markow DB (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology, 29(3), 257–302. 10.1006/COGP.1995.1016 [PubMed: 8556847]

Wickham H (2016). Ggplot2: Elegant graphics for data analysis. New York, NY: Springer-Verlag Retrieved from https://ggplot2.tidyverse.org

Wickham H (2017). Tidyverse: Easily install and load the 'tidyverse'. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham H (2019a). Forcats: Tools for working with categorical variables (factors). Retrieved from https://CRAN.R-project.org/package=forcats

Wickham H (2019b). Stringr: Simple, consistent wrappers for common string operations. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham H, & Bryan J (2019). Readxl: Read excel files. Retrieved from https://CRAN.R-project.org/package=readxl

Wickham H, François R, Henry L, & Müller K (2019). Dplyr: A grammar of data manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham H, & Henry L (2019). Tidyr: Tidy messy data. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham H, Hester J, & Francois R (2018). Readr: Read rectangular text data. Retrieved from https://CRAN.R-project.org/package=readr

Xie Y (2015). Dynamic documents with R and knitr (2nd ed.). Boca Raton, FL: Chapman; Hall/CRC Retrieved from https://yihui.name/knitr/

Zhu H (2019). KableExtra: Construct complex table with 'kable' and pipe syntax. Retrieved from https://CRAN.R-project.org/package=kableExtra
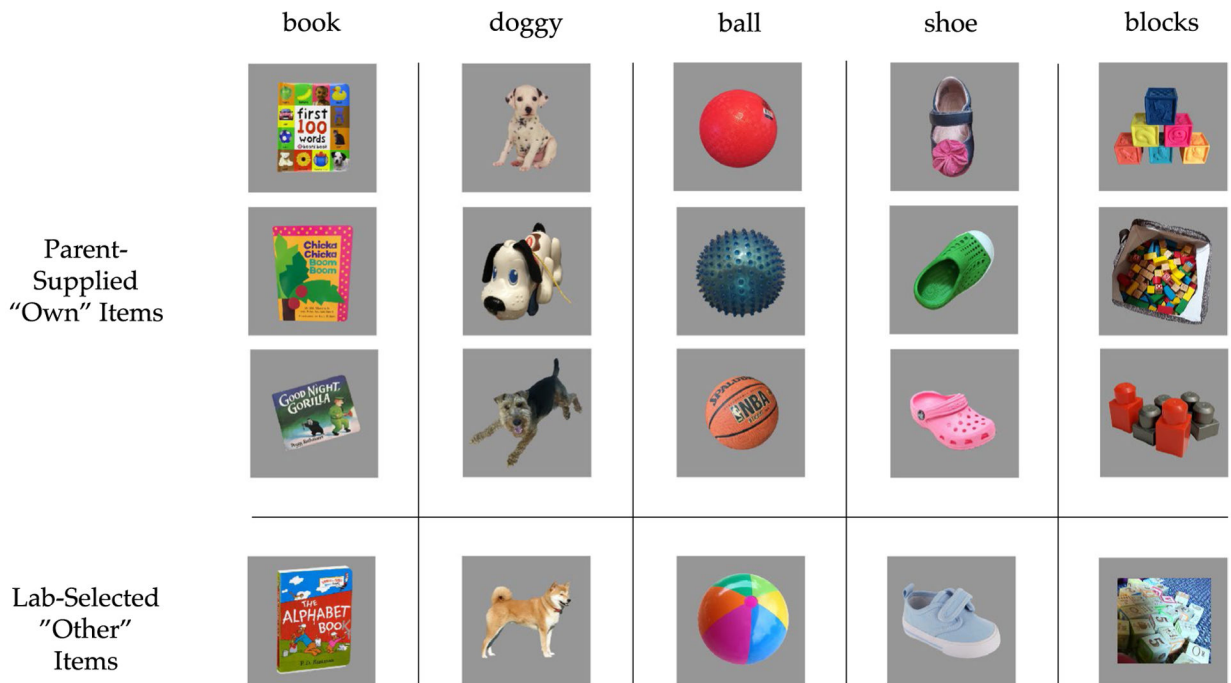
FIGURE 1.
Examples of a parent-selected "own" and laboratory-selected "other" items tested. Parents provided photographs based on a list of 26 "suggested objects." Samples of the five most frequently received and tested object are shown above. Images of all items used can be found on OSF (see text for link)
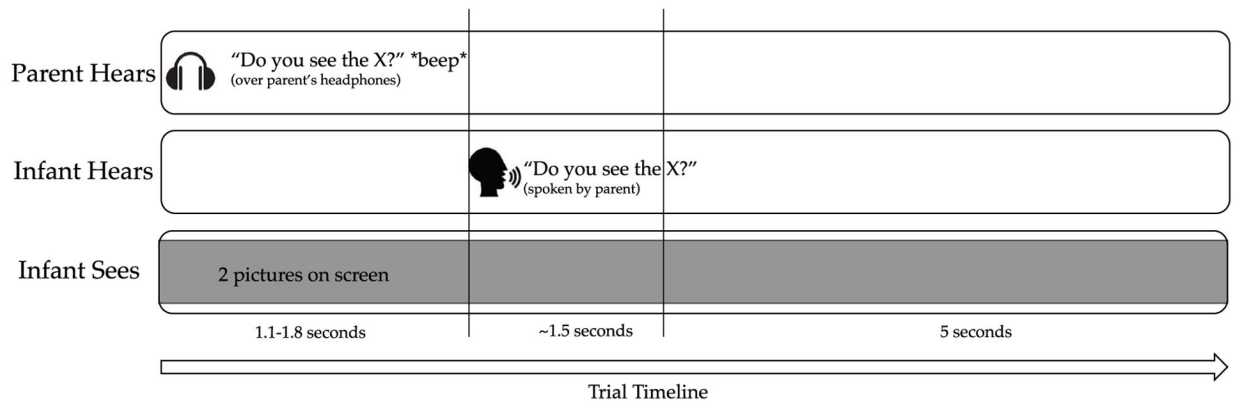
FIGURE 2.

Sequence of one test trial: Parents heard test sentences over headphones, followed by a beep. After the beep, they repeated them aloud to their infant. The images were displayed from trial start until 5 s after the beginning of the target word (as said by the parent). Representative item pairs for own-item and other-item trial images for a single item pair are also shown below the timeline
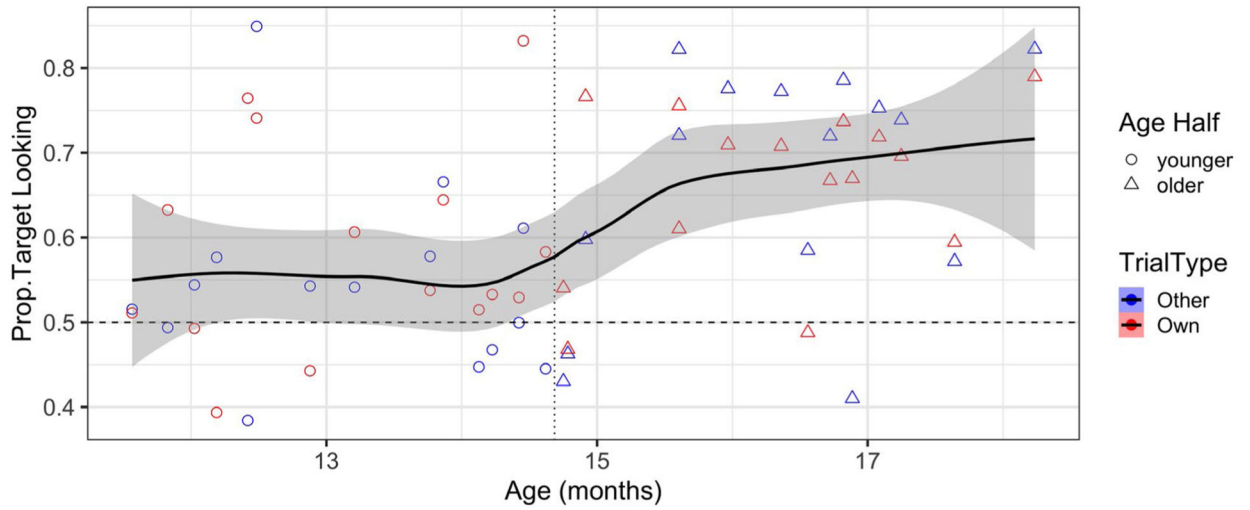
FIGURE 3.

Mean proportion of target looking over the 367- to 2,000-ms window of analysis, by age and trial type (other vs. own). Individual infants' trial-type means are shown by each point (older = triangles, younger = circles). Chance looking (0.5) is indicated by a dashed horizontal line, and median age is represented by a dotted vertical line. The curve and 95% CI are from a loess fit (i.e., local estimator); since trial type did not predict performance in any analysis, a single line is fit, though trial-type means for each infant are depicted for data transparency (other in blue, own in red)
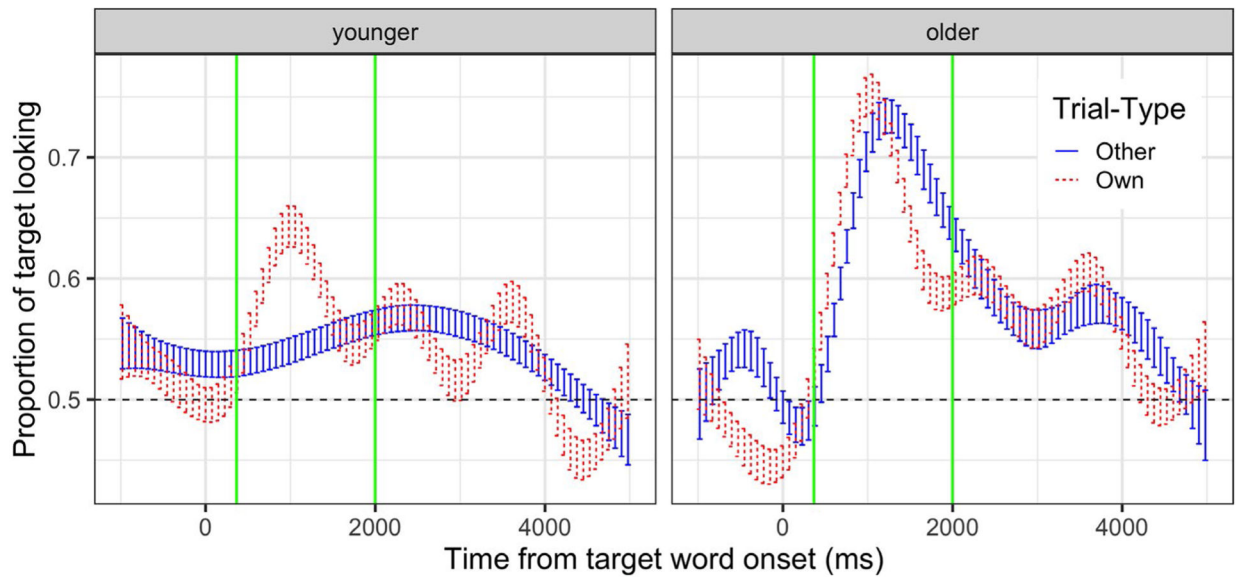
FIGURE 4.
Time course of mean proportion of target looking (aggregated by infant), split by median
age (14.7 m) and trial type (other vs. own). Time 0 is target word onset (e.g., "bottle" in
"Where's the bottle?"). The window of analysis (367–2,000 ms, as in Swingley & Aslin
2000) is indicated by vertical green lines. Chance looking (0.5) is indicated by a dashed
horizontal line

FIGURE 5.
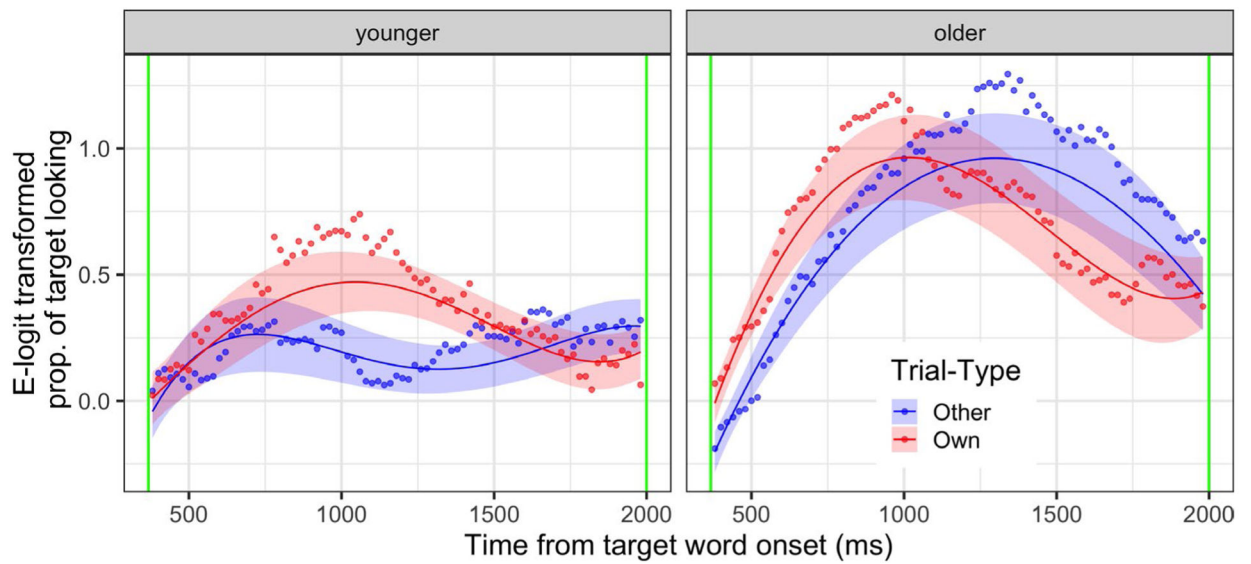Growth curve estimates of e(mpirical) logit-transformed mean proportion of target looking during analysis window, split by median age (14.7 m) and trial type (other vs. own). Line represents model estimates; ribbon represents standard error. Points represent mean value over 20-ms timebin, infant, and trial type. e-logit value of 0 corresponds to a proportions of 0.5 in the preceding figure

TABLE 1

Mean word exposure (range: 1–5) and CDI scores by age-half and over all infants

Age_half	M_exposure	CDI_U	CDI_Test_U	CDI_P	CDI_Test_P	Age
Older	4.52	218.93	4.20	44.27	3.00	497.53 (32.11)
Younger	4.27	96.27	4.80	6.13	0.67	401.93 (32.49)
All Ss	4.39	157.60	4.50	25.20	1.83	449.73 (58.06)

Note: Age is in days. Note that 19/87 items we tested across infants were not on the CDI (which has 396 vocabulary items), for example, giraffe.

Abbreviations: CDI, Communicative Development Inventory; P, produced; Test, tested items from study (8 per child); U, understood.

TABLE 2

Multi-level regression results

	Dependent variable		
	Mean proportion of target looking		
	(1)	**(2)**	**(3)**
Trial type		0.018 (0.023)	0.018 (0.023)
Centered age	0.029 *** (0.010)	0.029 *** (0.010)	0.033 *** (0.011)
Trial type: age			−0.008 (0.012)
Intercept	0.613 *** (0.018)	0.604 *** (0.021)	0.604 *** (0.022)
Observations	60	60	60
Log likelihood	40.903	38.359	35.078
Akaike Inf. Crit.	−73.807	−66.717	−58.155
Bayesian Inf. Crit.	−65.430	−56.246	−45.589

Note: Random effects for all models are (1|Subject).

*
$p < .1$;

**
$p < .05$; and

***
$p < .01$.

TABLE 3

Proportion of target looking by age-group and trial type (Comp. = comparison; chance = 0.5)

Age-group	Trial type	M (SD)	Comp. to chance	t test p-val[a]	#subjs > chance	Binomial test p-val
Older	Other	0.664 (0.143)	$t(14) = 4.442$	<.001	12/15	.035
Older	Own	0.661 (0.1)	$t(14) = 6.234$	<.001	13/15	.007
Younger	Other	0.544 (0.11)	$t(14) = 1.545$	.145	9/15	.607
Younger	Own	0.584 (0.122)	$t(14) = 2.665$	.018	12/15	.035

[a] *t* test *p*-values above are from the more conservative two-tailed test and are presented uncorrected. We note that applying a Holm–Bonferroni correction does not change the pattern of significance of these results.

TABLE 4

Fixed effects for growth curve model, younger infants

Term	Estimate	SE	T	df	p-value
(Intercept)	0.19	0.09	2.13	15.20	.05
TrialTypeOwnItem	0.11	0.02	7.01	2,358.54	<.001
Time1	0.08	0.19	0.42	15.56	.682
Time2	0.04	0.10	0.39	17.17	.701
Time3	0.18	0.09	2.10	18.13	.05
Time4	−0.12	0.07	−1.62	20.13	.121
TrialTypeOwnItem:time1	−0.14	0.05	−2.61	2,362.43	.009
TrialTypeOwnItem:time2	−0.44	0.05	−8.33	2,365.87	<.001
TrialTypeOwnItem:time3	−0.01	0.05	−0.25	2,372.83	.806
TrialTypeOwnItem:time4	0.18	0.05	3.53	2,367.53	<.001

Note: Random effects for all models are (1 + time1 + time2 + time3 + time4 | SubjectNumber; see text for details).

TABLE 5

Fixed effects for growth curve model, older infants

Term	Estimate	SE	T	df	p-value
(Intercept)	0.66	0.12	5.66	15.18	<.001
TrialTypeOwnItem	0.00	0.02	−0.26	2,362.12	.795
Time1	0.57	0.29	1.96	15.24	.069
Time2	−0.87	0.19	−4.60	15.80	<.001
Time3	0.04	0.12	0.32	16.54	.751
Time4	0.00	0.09	−0.01	18.90	.991
TrialTypeOwnItem:time1	−0.63	0.06	−11.01	2,364.02	<.001
TrialTypeOwnItem:time2	0.11	0.06	1.94	2,372.29	.053
TrialTypeOwnItem:time3	0.31	0.06	5.36	2,373.14	<.001
TrialTypeOwnItem:time4	0.04	0.06	0.76	2,370.75	.447

Note: Random effects for all models are (1 + time1 + time2 + time3 + time4 | SubjectNumber; see text for details).

TABLE 6

Mean adult similarity scores (range:1–7) and infant word exposure (range:1–5) for common target words (>5 infants)

Target	Adult_mean_sim	Mean_exposure	*n*_kids
Book	2.52 (1.51)	4.91 (0.29)	22
Ball	3.27 (1.43)	4.65 (0.7)	17
Shoe	2.95 (1.59)	4.4 (0.7)	10
Bottle	2.24 (1.19)	4.62 (0.52)	8
Highchair	2.91 (1.36)	4.71 (0.49)	7
Blocks	3.15 (1.48)	4 (1.1)	6
Car	4.24 (1.52)	4.5 (0.84)	6
Diaper	2.52 (1.31)	5 (0)	6
Spoon	2.36 (1.3)	3.83 (1.17)	6

Note: Full data for each item tested available on OSF.