Research Article

An Eye-Tracking Study of Receptive Verb Knowledge in Toddlers

Matthew James Valleau,[a] Haruka Konishi,[b] Roberta Michnick Golinkoff,[c]
Kathy Hirsh-Pasek,[d] and Sudha Arunachalam[a]

Purpose: We examined receptive verb knowledge in 22- to 24-month-old toddlers with a dynamic video eye-tracking test. The primary goal of the study was to examine the utility of eye-gaze measures that are commonly used to study noun knowledge for studying verb knowledge.

Method: Forty typically developing toddlers participated. They viewed 2 videos side by side (e.g., girl clapping, same girl stretching) and were asked to find one of them (e.g., "Where is she clapping?"). Their eye-gaze, recorded by a Tobii T60XL eye-tracking system, was analyzed as a measure of their knowledge of the verb meanings. Noun trials were included as controls. We examined correlations between eye-gaze measures and score on the MacArthur–Bates Communicative Development Inventories (CDI; Fenson et al., 1994), a standard parent report measure of expressive vocabulary to see how well various eye-gaze measures predicted CDI score.

Results: A common measure of knowledge—a 15% increase in looking time to the target video from a baseline phase to the test phase—did correlate with CDI score but operationalized differently for verbs than for nouns. A 2nd common measure, latency of 1st look to the target, correlated with CDI score for nouns, as in previous work, but did not for verbs. A 3rd measure, fixation density, correlated for both nouns and verbs, although the correlation went in different directions.

Conclusions: The dynamic nature of videos depicting verb knowledge results in differences in eye-gaze as compared to static images depicting nouns. An eye-tracking assessment of verb knowledge is worthwhile to develop. However, the particular dependent measures used may be different than those used for static images and nouns.

Understanding how toddlers acquire the meanings of verbs has been a central endeavor of language acquisition research in recent decades. Because of their semantic and syntactic properties, learning verbs spurs the acquisition of more complex language. Semantically, verbs can denote events, allowing toddlers access to a rich repertoire of event meanings and relations among people and objects. Syntactically, verbs determine, to a large extent, what other elements can appear in a sentence and in what structural relationships. Adding verbs to the lexicon can thus have cascading effects on language development (e.g., Goldfield & Reznick, 1990; Konishi, Stahl,

Golinkoff, & Hirsh-Pasek, 2016; Marchman & Bates, 1994). Indeed, recent evidence indicates that toddlers' production of verbs at 24 months of age predicts their later grammar skills better than their production of nouns (Hadley, Rispoli, & Hsu, 2016).

However, these critical elements of the lexicon are notoriously difficult to acquire for English-learning toddlers. The first words they produce are typically names for people and objects and words that are part of social routines (such as "hi"); verbs are rare in the productive lexicon until well into the second year of life (e.g., Fenson et al., 1994; Naigles, Hoff, & Vear, 2009). This pattern is mirrored in comprehension: Bergelson and Swingley (2012) found that even 6-month-olds know the meanings of some common nouns (see also Tincoff & Jusczyk, 1999, 2012), but only at 10 months of age do they seem to know the meanings of any verbs at all (Bergelson & Swingley, 2013). Laboratory studies introducing novel words also reveal that verb learning requires more supportive learning situations than noun learning (see Gentner, 2006; L. R. Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005, for reviews). In particular, supportive learning situations for acquiring verb meanings typically include informative syntactic contexts,

[a]Department of Speech, Language, & Hearing Sciences, Boston University, MA
[b]Department of Human Development and Family Studies, Michigan State University, East Lansing
[c]University of Delaware School of Education, Newark
[d]Department of Psychology, Temple University, Philadelphia, PA
Correspondence to Sudha Arunachalam: sarunach@bu.edu

such as sentences that reveal the number and types of event participants with which the verb's referent event can occur (e.g., Fisher, 1996, 2002; L. Gleitman, 1990; Naigles, 1990; Yuan & Fisher, 2009). This presents a puzzle: Syntactic information supports the acquisition of verb meaning even as verb knowledge supports the acquisition of syntax. Verb knowledge and syntactic knowledge must develop in tandem, each informing the other and together influencing the trajectory of language development (e.g., Gertner, Fisher, & Eisengart, 2006; L. R. Gleitman et al., 2005; Golinkoff & Hirsh-Pasek, 2008; Hadley et al., 2016; He & Arunachalam, 2017; Trueswell & Gleitman, 2007).

Verb acquisition thus plays a multifaceted role in early language development, presenting challenges as well as rich rewards for learners. Therefore, understanding precisely what toddlers' early verb lexicons look like in the first years of life may have enormous benefits. Although verbs are challenging even in typical development, children with language disorders often have particular difficulty with verbs (e.g., Hadley, 2006; Olswang, Long, & Fletcher, 1997), and the ability to evaluate their verb knowledge and intervene early may impact their language outcomes. A better understanding of early verb knowledge will also shape theories of language development by establishing how syntactic and lexical knowledge interact. Many studies have investigated early syntactic understanding, revealing that toddlers have more competence in comprehension than in production. For example, by 17 months of age, toddlers use word order to identify the thematic relationships between the event participants named in a sentence (Hirsh-Pasek & Golinkoff, 1996); by 21 months of age, they do so even with novel verbs (Gertner et al., 2006). By 24 months of age, toddlers build up syntactic representations for sentences they hear incrementally, as the sentences unfold (Bernal, Dehaene-Lambertz, Millotte, & Christophe, 2010). The other side of this picture—how many verbs toddlers know when they begin to comprehend basic syntactic structures and how their verb vocabularies grow in tandem with syntactic development—is less well understood. To flesh out this picture, we must have a better understanding of toddlers' receptive verb vocabularies, given that receptive knowledge generally precedes expressive use.

However, receptive vocabulary is particularly challenging to measure at young ages (e.g., Friend & Keplinger, 2003), and verb knowledge may be even more difficult to study than noun knowledge. Parent report of vocabulary is the most common assessment method in young children, but parent report is likely to provide a poor estimate of verb knowledge (Houston-Price, Mather, & Sakkalou, 2007; Tardif, Gelman, & Xu, 1999). Parent report may be influenced by, for example, toddlers' abilities to respond appropriately in the course of a routine or an activity rather than their understanding of the meanings of the verbs (Tomasello & Mervis, 1994). Further, parent report of receptive vocabulary knowledge is only valid for young toddlers, before they produce many verbs; the authors of a widely used parent report checklist, the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1994), solicit parent report of expressive knowledge rather than receptive knowledge for toddlers over 16 months of age (see also Rescorla, 1989), reasoning that parents are not able to keep track of receptive vocabulary once toddlers have begun talking.

Direct assessment techniques may be more useful for measuring receptive vocabulary, but they also have limitations. Assessments such as the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) require children to point to pictures of objects or actions and thus are only appropriate for ages 2.5 years and older. Further, these assessments depict verb meanings with static drawings, and young children have difficulty inferring motion from these (Cocking & McHale, 1981; Friedman & Stevenson, 1975)—they may thus underestimate verb knowledge.

An excellent measure for studying language comprehension in childhood is eye-gaze: It does not require verbal or gestural response and is suitable even for young infants and for children with communication disorders. Eye-tracking studies using two related paradigms, the intermodal preferential looking paradigm (IPLP; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1996) and looking-while-listening (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Fernald, Zangl, Portillo, & Marchman, 2008), have offered the field a window into toddlers' language abilities (see Golinkoff, Ma, Song, & Hirsh-Pasek, 2013, for a recent review). Recent work asked whether these paradigms can be used as an alternate technique for assessing vocabulary knowledge: Houston-Price et al. (2007), Killing and Bishop (2008), and Styles and Plunkett (2009) tested eye-tracking versions of the imageable concrete nouns on the Oxford CDI, finding that eye-gaze measures are indeed excellent for studying young children's noun vocabularies (see also Reznick, 1990).

Although many studies have used eye-gaze to study receptive noun knowledge, fewer have included verbs. A major advantage of eye-tracking approaches for verbs is that it is easy to present videos to depict dynamic event referents, instead of line drawings or still photographs. Huttenlocher, Smiley, and Charney (1983) were the first to test receptive verb knowledge using dynamic video scenes, although their measure was pointing rather than eye-gaze. They tested 20 verbs and found that children between 22 and 42 months of age understood movement verbs (e.g., *jump*) better than verbs of change of state or location (e.g., *bring*). Golinkoff et al. (1987) also used eye-gaze measures to study 16-month-old toddlers' comprehension of 12 verbs referring to dynamic video scenes. Overall, they found that toddlers both oriented faster to and looked longer at the matching than the nonmatching scene. Naigles (1997) found that 17.5-month-olds only knew one of six tested verbs (*roll*), although by 27 months of age, they showed evidence of knowing all six. Naigles and Hoff (2006), however, tested 17-month-olds with 12 verbs, finding relatively strong knowledge overall. The youngest infants tested thus far have been those aged 6–16 months: Bergelson and Swingley (2013) reported that,

by 10 months of age, infants showed some evidence of understanding "abstract" words, which included some verbs. In all of these studies, the fact that children showed evidence of knowing some verb meanings, despite parental report of few or no verbs in their expressive vocabularies, suggests that eye-gaze measures can be particularly valuable for tapping into receptive verb knowledge at young ages. With older children, aged 3–6 years, Goldfield, Gencarella, and Fornari (2016) also found that increased looking time from a baseline period to a test period with dynamic scene stimuli served as a good measure of verb knowledge and that this measure correlated with children's PPVT scores.

This brief overview of the literature reveals that, although eye tracking is a promising technique for studying receptive vocabulary, including verbs, it also raises methodological questions. Eye tracking offers a rich data set; many measures can be extracted to examine different aspects of toddlers' knowledge and processing. Previous research using nouns and static images has primarily focused on two measures: accuracy, or how much time is spent looking at the target image (as compared to the distractor), and latency, or how quickly children direct their first look to the target image. These measures correlate with children's (earlier and later) vocabulary score on the CDI (Fernald, Perfors, & Marchman, 2006; Marchman & Fernald, 2008) and are thus used in numerous studies as indicators of individual children's knowledge. However, for verbs and dynamic scenes, these eye-gaze measures may play out differently— after all, dynamic scenes capture attention continually as they move. Most of the eye-tracking work described above that has studied verb knowledge with dynamic scenes uses accuracy as a measure (although accuracy has been operationalized in slightly different ways), and a handful of studies have explored other measures as well.

The goal of the current line of work is to investigate toddlers' verb knowledge using measures of their eye-gaze as they view dynamic video scenes—but this study focuses on the prerequisite methodological issue. We investigate multiple eye-gaze measures to ask: Which measures are useful for assessing individual differences in receptive verb knowledge?

We focus on toddlers at 22–24 months of age who have only recently begun producing their first verbs and two- to three-word combinations; they are undoubtedly actively engaged in the process of adding new verbs to their receptive vocabularies but are still producing relatively few (e.g., Naigles et al., 2009). At this age, assessment of vocabulary knowledge, and particularly verb knowledge, using standardized child assessments is difficult, given that toddlers often do not point reliably to videos in response to a prompt (e.g., Arunachalam & Waxman, 2011). Nevertheless, this is perhaps the most important age group to study, given that we are just beginning to see evidence of syntactic comprehension and incremental structure building (e.g., Bernal et al., 2010; Hirsh-Pasek & Golinkoff, 1996) and that this is when toddlers are typically identified as late talkers (e.g., Rescorla & Dale, 2013).

We adapted a verb comprehension task recently developed by Konishi et al. (2016). In their study, children aged 27–34 months (M_{age} = 30 months) were shown two dynamic video scenes depicting common verb referents side by side (e.g., girl clapping, same girl stretching) and were asked to point to the one depicting a particular verb label (e.g., "Where is she clapping?"). They found that, on average, children pointed correctly 80% of the time, and their performance on this task correlated with their score on the PPVT, suggesting that this was an appropriate measure of individual differences in verb knowledge. To adapt this assessment for slightly younger ages—those who are not old enough to provide systematic pointing responses—we used Konishi et al.'s stimuli to create an eye-tracking task. We expect that, as a group, the receptive verb vocabularies of the near 2-year-olds we study will be smaller than the 80% of the tested verbs known by Konishi et al.'s 30-month-olds (but, too, that they will know more than the handful of verbs known by 10-month-olds in Bergelson and Swingley, 2013). However, the pressing question is whether we can characterize individual toddlers' vocabularies relative to other toddlers. This would ultimately be useful for identifying those who may require intervention.

To begin to address this question, in this study, we examine three eye-gaze measures and look for correlations between these measures and performance on the CDI (Words and Sentences, Short Form A; Fenson et al., 1993) as evidence that a particular eye-gaze measure is a good reflection of toddlers' vocabulary knowledge. Note that the CDI is a measure of expressive knowledge, rather than receptive knowledge, but this is the standard measure for parent checklists at this age. Although we have said that parent report checklists such as the CDI may incorrectly estimate verb knowledge, parent report is currently the best available tool for assessing expressive vocabulary knowledge in this age group.

Given differences between the dynamic scenes we use to depict verbs as compared to the static images used to depict nouns in prior work, we predict that there will be different relationships between parent report and eye-gaze measures between nouns and verbs. In addition to examining the standard two measures of accuracy and latency, we added a third measure, fixation density, based on research from the literature on how adults examine complex visual scenes (e.g., Henderson, Weeks, & Hollingworth, 1999), reasoning that our dynamic scenes are similarly complex and less like the static images used to depict noun referents. Thus, we tested three eye-gaze measures—accuracy, latency, and fixation density—described in more detail below, to ascertain which measure(s) might lend insight into individual toddlers' verb knowledge.

Method

Participants

Forty toddlers (20 girls and 20 boys) between the ages of 22.0 and 24.9 months (M = 23.0, SD = 0.7) were

recruited from the greater Boston area and included in the final sample. Toddlers were reported by their parents to be full-term, typically developing, and raised in English-speaking environments with no more than 30% exposure to other languages. One toddler was reported by his or her parent to be Asian, eight were reported to be multiracial (one Hispanic), and 30 were reported to be White (two Hispanic); no response was provided for the remaining child. The highest education level among the parents was reported to be a graduate degree for 24 toddlers, a bachelor's degree for nine, an associate's degree for one, and a high school diploma for one; four parents did not provide a response. Annual household income was reported as ≥ $150,000 for eight toddlers, between $100,000 and $149,999 for 10 toddlers, between $80,999 and $99,999 for three toddlers, between $60,999 and 79,999 for six toddlers, between $40,999 and $59,999 for one toddler, between $20,000 and $39,999 for one toddler, and under $20,000 for one toddler; 10 parents did not provide a response. Seven additional toddlers were tested but excluded from analysis due to either poor eye tracking (less than 50% tracking over the course of the session; $n = 5$), fussiness ($n = 1$), or parental interference ($n = 1$). Two toddlers were included who did not complete the entire task, one who contributed data from 80% of trials and the other from 52%.

Toddlers were assigned to one of four experimental lists (A, B, C, or D), described below; each had equal numbers of boys and girls, and they did not differ in age or score on the CDI (Words and Sentences, Short Form A; Fenson et al., 1993). In Table 1, we report the participants' ages and CDI scores within each of the four lists. We report their CDI scores for all of the words on the list as well as for the subset of the list containing nouns ($n = 50$) and the subset containing verbs ($n = 14$). The CDI norms provide information about the percentile range into which scores fall by age (in months) and sex; the mean of percentiles across the sample was 53rd, and only two toddlers in the sample were in the 10th percentile or below. All toddlers were reported to be combining words "sometimes" or "often."

Apparatus

Stimuli were presented on a 24-in. Tobii T60 XL eye tracker with built-in speakers and tracking accuracy of 0.5°, which captures gaze position approximately every 17 ms. Toddlers sat either in a car seat approximately 50–80 cm from the eye tracker or on the parent's lap (if the latter, the parent wore a blindfold).

Stimuli and Design

We adapted stimuli from Konishi et al. (2016). Konishi et al. selected 36 early-acquired verbs, 75% of which appear on the CDI and the others of which are of comparable frequency in child-directed speech (according to the Child Language Data Exchange System Parental Corpus; Li, 2001) and have been used in other studies of early vocabulary (Masterson, Druks, & Galliene, 2008). However, a critical inclusion criterion for verbs in their study (and therefore in ours) was that the verbs must label highly imageable concepts that can be easily depicted in videos. They also chose 14 nouns, some of which they used as warm-up trials; we included all 14 nouns in the current study. See Konishi et al. for more details about stimuli selection (see Table 2).

Visual Stimuli

We used Konishi et al.'s visual stimuli, in which verb trials depicted two dynamic video recordings of actors and objects side by side and noun trials depicted two still photographs of objects and animals, as shown in Figure 1. For verb trials, eight of the 36 videos involved one animate event participant (e.g., girl clapping), whereas the other 28 videos involved two event participants, one animate and the other inanimate (e.g., girl bouncing a ball). The videos were paired such that the same event participant(s) were depicted in the two videos for each trial (e.g., On one trial, one scene depicted a girl clapping, and its pair depicted the same girl stretching).

Auditory Stimuli

A female native speaker of English recorded sentences in a sound-attenuated booth using a child-directed

Table 1. Participants' mean (*SD*) ages and scores on the MacArthur-Bates Communicative Development Inventories (CDI): Words and Sentences, Short Form A (a measure of expressive vocabulary), by Lists A–D.

	A	B	C	D
N	10	10	10	10
Age, months	23.2 (0.7)	23.3 (0.8)	22.8 (0.7)	22.7 (0.4)
CDI total	54.7 (17.6)	53.6 (16.2)	57.8 (29.4)	59.1 (27.3)
CDI nouns subset	32.4 (9.6)	31.6 (10.1)	31.6 (14.8)	30.7 (13.1)
CDI verbs subset	4.4 (3.6)	4.2 (2.7)	5.2 (5.3)	6.0 (6.2)

Note. Lists reflect assignment to one of four experimental lists differing in which word of the pair is queried and in trial order.

Table 2. List of nouns and verbs included as stimuli.

Nouns		Verbs	
Airplane	Bite	Drink	March
Orange	Drop	Pour	Spin
Banana	Blow	Eat	Open
Cookie	Squeeze	Push	Shake
Bird	Bounce	Feed	Read
Fire truck	Roll	Hug	Rip
Crab	Break	Jump	Rock
Pancakes	Lick	Run	Wash
Doughnut	Clap	Kick	
Goldfish	Stretch	Throw	
Giraffe	Cry	Kiss	
Rocket ship	Dance	Tickle	
Grapes	Cut	Lift	
Squirrel	Tie	Pull	

Note. The two words depicted as a pair in the visual stimuli are listed together.

Figure 1. Structure of one representative noun trial and one representative verb trial, arrayed from left to right.

Phase	INSPECTION PHASE		BASELINE PHASE	QUERY PHASE	RESPONSE PHASE
NOUN TRIAL					
Visual Stimuli	[cookie]	[banana]	[cookie and banana]	[star]	[cookie and banana]
Duration	2 s	2 s	3 s	4 s	6 s
Auditory Stimuli	Oh, look!	Wow!	Do you see?	Where is the cookie?	Find the cookie! Do you see the cookie?
VERB TRIAL					
Visual Stimuli	[woman clapping]	[woman clapping]	[two women]	[star]	[two women]
Duration	4 s	4 s	6 s	4 s	6 s
Auditory Stimuli	Oh, look!	Wow!	Do you see?	Where is she clapping?	Find clapping! Do you see where she's clapping?

speech register. The audio for noun trials consisted of simple carrier phrases (e.g., "Where is the *cookie*?"), whereas for verb trials, it varied by whether the depicted event involved one or two participants; one-participant events were queried with intransitive syntax (e.g., "Where is she *clapping*?"), and two-participant events were queried with transitive syntax (e.g., "Where is she *bouncing* the ball?"). The carrier frames varied in this way because we used what we thought was the most discourse-natural query for the different event types. For both event types, we also included queries in neutral syntax (e.g., "Find *clapping*!"). Because the same event participants were depicted in each video in a pair, the verb served as the only cue to which scene was the target.

We synchronized the auditory and visual stimuli according to the timeline depicted in Figure 1. On each trial, during the inspection phase, each dynamic scene or still image was first presented individually, followed by a baseline phase where they appeared simultaneously side by side, accompanied by language designed to draw toddlers' attention to the screen (e.g., "Wow! Do you see?"). Their gaze during this phase served as a measure of the scenes' baseline salience. Next, during the query phase, a white screen with a centrally placed star appeared, along with the test query designed to draw attention to the target

(e.g., "Where is she clapping?"). We included the centrally placed fixation star so that toddlers could hear the test query without distraction from the visual scenes and so that their gaze would be centrally located before the scenes reappeared on either side of the screen (see also Delle Luche, Durrant, Poltrock, & Floccia, 2015). Finally, during the response phase, the two scenes or images reappeared in their original locations with two presentations of a similar test query (e.g., "Do you see where she's clapping? Find clapping!"). Noun and verb trials were structured identically except that noun trials were shorter in duration in the inspection and baseline phases; we chose this shorter duration because the noun stimuli were static, and we were worried that toddlers would be bored if they remained on the screen for too long.

The side presented first during the inspection phase (right vs. left) and which side was the target (right vs. left) were counterbalanced across trials. Toddlers were assigned to one of four lists. Lists A and B were identical except that they queried different members of the paired stimuli. For example, on the trial depicting clapping on one side of the screen and stretching on the other, List A queried "clap" and List B queried "stretch." We chose this between-subjects design because a within-subject design may have shown inflated vocabulary knowledge if toddlers used process of

elimination or mutual exclusivity and would also have made the task too long. Although the between-subjects design does not account for the fact that the two words within a pair may not be perfectly matched in difficulty, we controlled for this as best we could by carefully considering both the properties of the two words in a pair (e.g., both were either transitive or intransitive) and the characteristics of the toddlers assigned to each list (they did not differ in age, sex, or CDI score). Thus, 36 verbs were included, but each toddler was tested on 18 verbs. Lists C and D were identical to Lists A and B, respectively, except that the trials were presented in reverse order. Preliminary analyses indicated no effects of list; we collapsed across lists in the reported analyses.

Procedure

The procedure was approved by Boston University's Institutional Review Board. Toddlers played with an experimenter while parents provided written consent and completed the CDI. The parent and toddler were then brought into the testing room. One experimenter remained with the toddler, whereas another operated the eye tracker from behind a curtain.

After a 5-point calibration procedure (Tobii Studio 3.1.2), toddlers viewed two warm-up trials. On each trial, two video clips of Sesame Street characters were presented on the screen side by side, with a test query directing attention to one of them (e.g., "Where's Big Bird?"). These trials were intended to familiarize toddlers with the trial structure—that two scenes would appear simultaneously and they would be asked about one of them. We did not analyze gaze behavior during these trials. Then, toddlers participated in the task proper, consisting of 18 verb trials and seven noun trials, interspersed.

Measures and Analyses

We extracted the raw data from Tobii Studio and used them to study verb knowledge in several ways. To prepare the data, we focused only on two phases: the baseline phase and the response phase. We removed all data points for which there was track loss (e.g., eye blinks), which comprised 16% of the data,[1] and all data points for which toddlers were looking at neither the target nor the distractor scene (7%). We first conducted some preliminary analyses to provide a manipulation check and to broadly characterize the data at a group level before turning to our primary goal of examining specific eye-gaze measures and their utility for understanding individual differences in verb vocabulary.

First, we conducted a manipulation check by examining whether toddlers preferred the target scene during the response phase as compared to during the baseline phase. Given that we expected toddlers to know at least some of the tested words, such a check ensures that they are responding to the prompt heard during the query phase. We analyzed the data from noun trials and verb trials separately, given the critical differences between the stimuli types (i.e., static vs. dynamic). These analyses involved a now-standard approach (Barr, 2008). To reduce the effect of eye movement–based dependencies (viz. that the eyes cannot shift from one location to the other instantly and thus gaze at one time point is not independent of gaze at the next), this approach involves first binning the data into 50-ms bins by aggregating over three consecutive frames. Then, we transformed the binned data using an empirical logit function, which has been found to be a useful transformation for gaze data (Barr, 2008). We then conducted a mixed-effects regression with the transformed data as the dependent variable and including a by-subject random intercept and slope for time (in seconds), a by-item random intercept, and fixed effects for time (in seconds), phase (baseline coded as −0.5, response coded as 0.5), sex (male coded as −0.5, female coded as 0.5), and interactions of phase and time, using maximum likelihood estimation and the *lmer* package (Version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.3.0; R Core Team, 2014). Model comparison was done using the drop1() function with chi-square tests to determine whether individual effects contributed significantly to the model. As a post hoc analysis, we repeated the analysis for verbs only, but including whether they were one- or two-participant trials. We also conducted simple *t* tests for individual verbs to provide a picture of which verbs were best known overall.

To address the primary methodological goal of identifying which eye-gaze measures are well suited to examine individual differences, we examined three specific eye-gaze measures (accuracy, latency, and fixation density) and correlated them with toddlers' CDI scores. As we have pointed out, the CDI may not be an ideal measure, particularly for verbs. However, we use it as a comparison to our eye-gaze measures because, quite simply, we need some established standard against which to interpret the results and correlating with CDI is a standard procedure for nouns (e.g., Fernald et al., 2006; Marchman & Fernald, 2008)—although CDI measures expressive rather than receptive vocabulary. It is also possible that the CDI is a suboptimal measure across the board, such that even if parent report underestimates expressive verb knowledge, it would do so relatively systematically across the lexicon, and thus CDI score would still correlate with the eye-gaze measures.

To fully explore the possibilities of these eye-gaze measures, we examined the data within two time windows: a short window of 300–1800 ms of both the baseline and response phases and a full window of 300–6000 ms (for the response and baseline phases on verb trials; the baseline phase on noun trials lasted only 3000 ms). We chose these

---

[1]We neither asked for nor encouraged pointing, but some toddlers did so spontaneously: Seven toddlers pointed on one or two trials, and another seven pointed on three to six trials (25 never pointed, and we do not have codable video for the remaining toddler). These toddlers tended to point at the end of the trial, after the target window of most of our analyses; nevertheless, when toddlers point, they often block the eye tracker, and this contributes to the track loss percentage.

windows because, for nouns and static images, the shorter 1500-ms window is commonly used (e.g., Fernald et al., 2008) but, for verb trials, a longer window may be sensible given that visual attention is likely to be distributed more across the dynamic scenes (as indeed suggested by our data). We also used a strict track loss criterion; we excluded trials from individual toddlers for which the track loss rate was more than 25%. Our rationale for doing so was that some measures, particularly latency and fixation density, would be easily degraded with high track loss.[2] The three measures are fully described below.

Accuracy

For the accuracy measure, we used a metric of a 15% increase in looking time to the target from baseline to response. A 15% increase has been judged in prior work as a substantial increase in looking preference, indicating receptive knowledge of the word: Reznick (1990), using data from 8- to 20-month-olds, examined a number of criteria for estimating noun knowledge from eye-gaze and found that a 15% increase in target looking from a time interval before the test query to after the test query was the best criterion to use, with high test–retest reliability and sensitivity to vocabulary increases with age (see also Killing & Bishop, 2008; Reznick & Goldfield, 1992). Goldfield et al. (2016) used this criterion with older children (aged 3–5 years) and, notably, tested verbs as well as nouns. Thus, this measure is useful for examining individual differences for nouns and static scenes in infants below the age range in the current study (Reznick, 1990) and preschoolers above the age range (Goldfield et al., 2016), as well as for verbs and dynamic scenes for the older preschoolers (Goldfield et al., 2016). Therefore, we calculated for each toddler the proportion of trials on which they showed at least a 15% increase from baseline to response and correlated this proportion with their CDI score. We conducted this correlation within both the short time window and the long time window, as described above.

Latency

Latency is often used to study speed of lexical processing with nouns and static images (e.g., Fernald, McRoberts, & Swingley, 2001) and was used by Golinkoff et al. (1987) for verbs and dynamic scenes. Golinkoff et al. found that latencies to the target scene versus the distractor were only significantly different on one of their six trials. This is perhaps not surprising given that dynamic properties of the scenes may capture toddlers' attention in unpredictable ways; their first look may reflect a motion-based attentional shift rather than a shift to identifying the named target. Thus, despite the robustness of this measure for nouns and static images, we suspected that this might not be the best measure for dynamic scenes.

We identified the first frame on which toddlers looked at the target during the response phase for at least 100 ms; we excluded any looks to the target within the first 50 ms, reasoning that extremely early looks would reflect chance looking at the onset of the response phase.[3] Correlations are expected to be negative because shorter latencies indicate faster performance. We only measured latency for the first look to the target scene, so the two time windows used for the accuracy analysis were not relevant.

Fixation Density

We were motivated to examine fixation data based on research about how adults examine complex visual scenes (Henderson et al., 1999), because our dynamic scenes on verb trials are more visually complex than the static images depicted on noun trials. Henderson et al. showed adults static but complex visual scenes (e.g., a bar) with either a semantically consistent element (e.g., a cocktail) or a semantically inconsistent element (e.g., a microscope) and found that adults had a higher proportion of fixations to the inconsistent than consistent element. We suspected that, similarly, toddlers might check back and inspect more carefully the scene that was inconsistent with the auditory stimuli, especially given that the visual scenes are dynamic—toddlers might need to verify that the depicted event continued to be different from the named event. Note that this "checking back" does not necessarily mean that toddlers do not prefer the target overall; these checking looks may be very brief. Thus, knowledge indicated by a large number of checking looks is consistent with knowledge indicated by overall preference for the target.

We computed the proportion of the number of fixations within the distractor scene to the number of fixations within the target scene. For example, for the contrived saccade pattern shown in Figure 2 for the target "clap," in which the circles represent fixations (with larger diameters indicating longer fixations), lines represent saccades, and numbers represent the order of the fixations, fixation density would be 1.5, with six fixations within the distractor scene and four within the target scene. Examining proportions rather than raw numbers of fixations allows us to control for the fact that some toddlers will simply look back and forth more than the others. This measure is therefore different from the number of times toddlers shift their attention from one scene to the other, which might be taken as a measure of uncertainty (Candan et al., 2012).

To obtain fixation data, we calculated the velocity of eye movements based on their relative position to the screen and the start and end points of the movement, using a velocity-threshold identification algorithm (Salvucci & Goldberg, 2000). Each movement was categorized as a

---

[2]The same patterns are evident if we include all trials, even those with > 25% track loss, although the correlations are weaker and, in some cases, not statistically significant.

[3]Note that Fernald et al.'s latency measure excludes the roughly 50% of trials on which toddlers happened to be looking at the target at the beginning of the analysis window. See Fernald et al. (2008) for discussion. However, in the current design, toddlers look at a central point before the scenes appear. Thus, we can include all trials.

Figure 2. A schematic illustration of fixations and saccades to the scenes in one verb trial; these hypothetical data are for illustration only. The circles represent fixations (with larger diameters indicating longer fixations), the lines represent saccades, and the numbers represent the temporal order of the fixations.
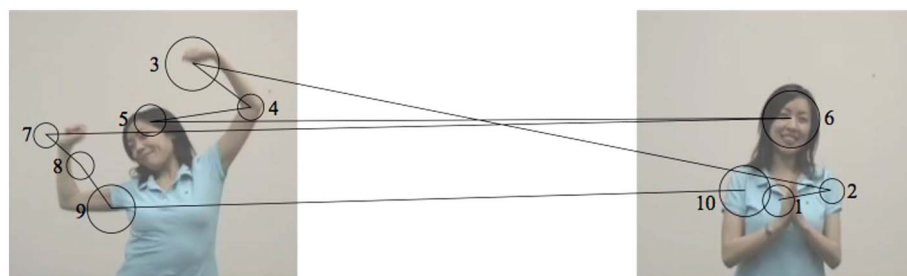
fixation if its velocity fell below a certain threshold (100 deg/s) or a saccade if it did not. Consecutive fixation points were collapsed, with the centroid calculated as the center of the fixation, and saccades were removed before analysis. We correlated toddlers' CDI score with fixation density in both the short and long time windows.

Results

Preliminary Analyses: Overall Gaze Patterns

We first present preliminary analyses to serve as a manipulation check and to characterize the data. First, we report descriptively on toddlers' overall looking patterns during the baseline and response phases for noun and verb trials and then report statistical analyses examining whether looking at the target increases significantly during the response phase. We then report on patterns for individual verbs.

During the baseline phase (3 s for nouns, 6 s for verbs), toddlers preferred to look at the target over the distractor 50% of the time for nouns and 50% of the time for verbs, indicating that toddlers show no preference for either scene before the test query. During the response phase (6 s for nouns and verbs), they preferred the target 72% for nouns and 64% for verbs, indicating that they shifted their attention to the target after the test query. Figure 3 depicts the time course of toddlers' gaze to the target scene on noun trials; and Figure 4, on verb trials, excluding the first 300 ms required for toddlers to program and launch an eye movement. (Recall that, just before the response phase, toddlers viewed a centrally placed star, so they had to launch an eye movement to gaze at either scene. The noise in the response phase lines at the very beginning of both figures reflects the fact that, on some trials, toddlers required more than 300 ms to shift their attention from where the central star had been to one of the scenes.) The y-axis indicates the proportion of frames on which toddlers looked at the target scene out of frames on which they looked at the target or distractor. Thus, values above 0.50 (the solid line) represent a preference for the target. Toddlers clearly preferred the target to the distractor during the response phase on both noun and verb trials, and they did so throughout the entire

window. However, the magnitude of the target preference during the response phases is steadier but smaller on verb trials than on noun trials, never reaching above about 0.65 (see Figure 4). This could be because their receptive verb knowledge is less robust and they therefore show a smaller preference for the target, but it could also be because dynamic scenes continually capture their attention such that, even as toddlers prefer the target, they continue to look back occasionally at the distractor scene. Although these two possibilities are not mutually exclusive, we will see evidence for the latter in the fixation density analysis below.

Because the noun and verb trials involved different stimulus types (static vs. dynamic), we did not combine them into a single analysis. We conducted two mixed-effects regression models, as described in the Analysis section, one for noun trials and the other for verb trials. These models were intended to assess whether the preference for the target was significantly greater in the response phase (after the test query) than the baseline phase (before the test query) and, further, whether there were effects of sex. Table 3 lists parameter estimates for the fixed effects for both the noun and verb models.

For nouns, the results indicate a significant effect of phase ($t = 22.15$, $p < .001$; toddlers prefer the target in the response phase as compared to the baseline phase) but no other main effects or interactions. For verbs, the results again indicate a significant effect of phase ($t = 13.13$, $p < .001$; toddlers prefer the target in the response phase as compared to the baseline phase), but also main effects of time ($t = 2.33$, $p < .03$) and interactions between time and phase ($t = 14.073$, $p < .001$) and between sex and phase ($t = -2.87$, $p < .01$). We do not interpret the latter given the lack of a main effect for sex (no crossover effect is evident, and the interaction is therefore likely spurious). The main effect of time and the interactions between time and phase reflect the strong and steady increase in looking at the target over time in the response phase, as compared to the steady "chance" looking in the baseline phase. We suspect that we did not see the main effect or interaction for nouns because, with nouns, children looked quickly and immediately at the target, but their target preference dropped over the course of the response phase. The lack of a main

Figure 3. Proportion of frames on which toddlers' gaze was directed to the target on noun trials, out of looks to either the target or the distractor, over the baseline and response phases. Error bars indicate standard error of subject means. Chance looking is 0.50 given that there are two images on the screen. The baseline phase was shorter in duration than the response phase on noun trials.
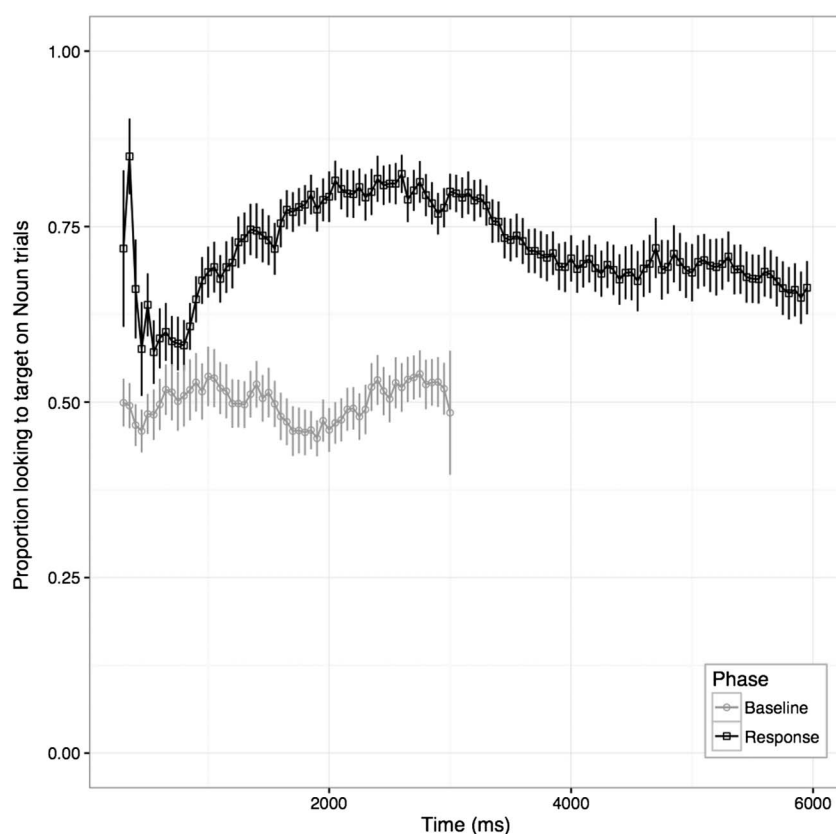

effect for sex for either nouns or verbs is interesting given that males' early vocabulary is generally lower than that of females (e.g., Fenson et al., 1994).

Among the verbs, recall that there were two different types of trials: eight depicted with only one event participant and labeled in intransitive sentences (e.g., "Where is she *jumping*?") and the other 28 depicted with two participants and labeled in transitive sentences (e.g., "Where is she *hugging* Cookie Monster?"). Toddlers show evidence of comprehending both intransitive and transitive syntactic structures, even with novel verbs, by the tested age (e.g., Arunachalam, Escovar, Hansen, & Waxman, 2013; Messenger, Yuan, & Fisher, 2015). However, there is also evidence that the more event participants that are involved in an event denoted by a verb, the more difficult the verb is to acquire—this has been found for typically developing children (e.g., Horvath, Rescorla, & Arunachalam, 2018), late talkers (Horvath, Rescorla, & Arunachalam, in press), and children with developmental language disorder (Thordardottir & Ellis Weismer, 2002). Further, the syntactic complexity of the test query differed in the two trial types. Therefore, to examine the potential confound of trial type, we repeated the mixed-effects

regression model as described above but including trial type as a fixed factor; this model yields no effect of trial type ($\beta = 0.062$, $t = 0.51$, $p = .61$), indicating that neither the number of event participants involved in the depicted events nor the syntactic complexity of the test query played a significant role in children's performance.

This overall pattern does not, of course, mean that toddlers knew all of the tested verbs. Figure 5 depicts toddlers' target preference for each individual verb during the baseline and response phases, again excluding the first 300 ms. The *y*-axis depicts frames on which toddlers looked at the target out of those frames on which they looked at either the target or distractor. Each verb is depicted next to the one with which it was paired in the visual stimuli. To assess statistically which verbs showed a reliable increase in preference from the baseline to response phases, we conducted a series of paired *t* tests on the mean difference between preference for the target scene during the baseline and response phases for each verb. With 36 verbs and, therefore, a Bonferroni-corrected α level of .0014, only four showed reliable increases: *blow, hug, roll,* and *tie.* Without correction and an α level of .05, *cry, dance, kick, rock, run, shake, spin, stretch,* and *wash* also showed

Figure 4. Proportion of frames on which toddlers' gaze was directed to the target on verb trials, out of looks to either the target or the distractor, over the baseline and response phases. Error bars indicate standard error of subject means. Chance looking is 0.50 given that there are two images on the screen.
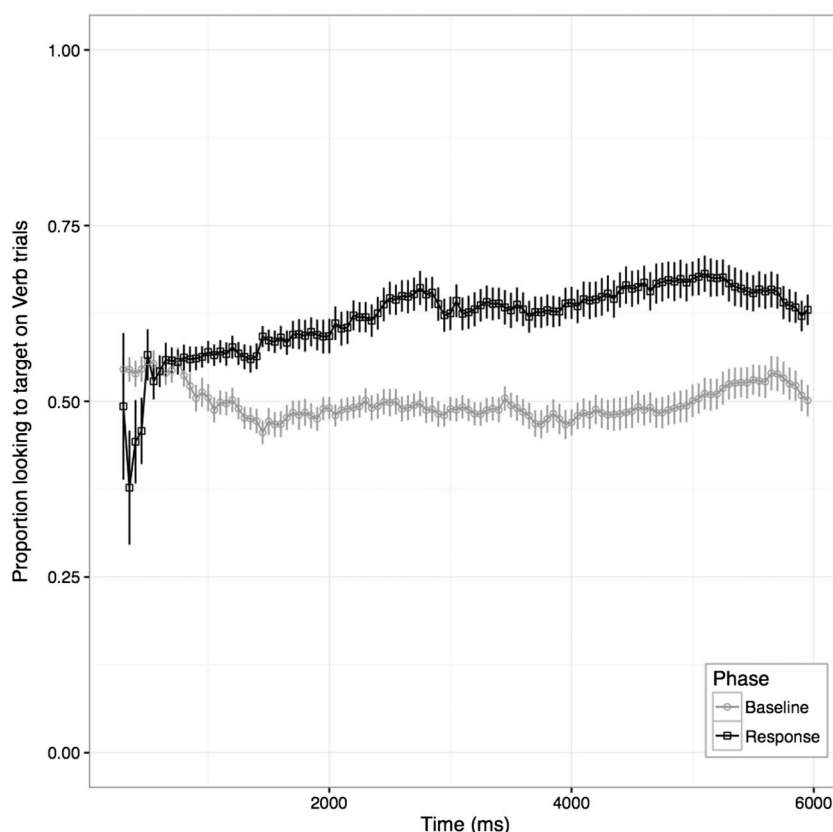
increases. Although a few of the verbs showed decreases in target preference from the baseline to response phases (e.g., *lift*), none of these was reliable.

Note that, for all of the verbs, except for *pull*, mean preference for the target during the response phase is above 0.50 (chance). Low performance with *pull* is unlikely to be attributable solely to the salience of its paired distractor scene (*lift*), because although toddlers showed a strong preference for the *lift* scene during the baseline phase, their preference for the *lift* scene when *lift* was queried was below the mean. It is more likely that both of these verbs are challenging for young learners or perhaps that the scenes depicting these verbs were insufficiently prototypical depictions of the events. However, on other trials, the particular distractor scene with which a target appeared is very likely to have affected performance. *Drop*, which was depicted with a very salient scene of a girl dropping a cupcake, yielded a much higher target preference than *bite*. Toddlers' attention may have been captured so strongly by the dropping event that, even when hearing *bite*, they continued to look at it even while they preferred the target (*biting*) scene overall. Relatedly, toddlers' preferences during the baseline phase are not independent of their preferences during the response phase. Toddlers may be at

"ceiling" in their preference for certain scenes, including the *drop* scene; the lack of change from the baseline to response phases may not necessarily indicate lack of knowledge of the verb's meaning.

Primary Analyses: Eye-Gaze Measures for Assessing Individual Differences

Next, to address the study's primary goal of examining multiple eye-gaze measures, we correlated toddlers' total vocabulary scores on the CDI with three different gaze measures: accuracy, which we operationalized as a 15% increase in looking time to the target from the baseline to response phases; latency of first look to the target; and fixation density, calculated as the proportion of the number of fixations to the distractor to the number of fixations to the target. For accuracy and fixation density, we used two time windows, 300–1800 and 300–6000 ms.

Accuracy

Building on prior work using a 15% increase in looking from the baseline to response phases (Goldfield et al., 2016; Reznick, 1990), we correlated the proportion of each toddler's trials showing this 15% increase with their CDI

Table 3. Parameter estimates from the best-fitting mixed-effects regression model of the effects of phase (baseline, response) and sex on the proportion of toddlers' gaze to the target scene (empirical logit transformed).

Model	Estimate	SD	t	p
Noun trials				
Intercept	0.29	0.072	4.13	
Time	−0.0038	0.019	−0.20	.84
Phase	−0.69	0.031	22.15	< .001*
Sex	−0.079	0.10	−0.75	.45
Time × Phase	−0.0080	0.017	−0.49	.65
Phase × Sex	−0.11	0.062	−1.71	.087
Time × Sex	0.0063	0.037	0.17	.86
Time × Phase × Sex	0.015	0.035	0.44	.66
Verb trials				
Intercept	0.12	0.057	1.80	
Time	0.036	0.015	2.33	.024*
Phase	0.22	0.017	13.13	< .001*
Sex	−0.054	0.062	−0.86	.39
Time × Phase	0.073	0.0052	14.073	< .001*
Phase × Sex	−0.097	0.034	−2.87	.0041*
Time × Sex	0.014	0.031	0.47	.64
Time × Phase × Sex	−0.0026	0.010	−0.26	.80

Note. The p values result from model comparison, indicating that the model is better fitting with that parameter than without.

*p < .05.

score, yielding the correlations shown in Table 4 (Measure 1). The results differ for the shorter and longer time windows. In the short window, nouns show a reliable correlation but there is no such correlation for verbs. In this window, the mean percentage of trials on which toddlers as a group showed this 15% increase for verbs was 43% (ranging from 12% to 69%, *SD* = 22%); for nouns, it was 55% (ranging from 0% to 100%, *SD* = 29%). The reverse holds for the longer time window; verbs show a reliable correlation, but nouns do not. Here, the mean percentage of trials on which toddlers showed the increase was 55% for verbs (14%–83%, *SD* = 14%) and 65% for nouns (0%–100%, *SD* = 26%).[4] Thus, the correlations suggest that the short window is better for nouns and the long window is better for verbs; the longer time window makes sense for verbs given the pattern in Figure 4, which suggests a protracted preference for the target over the course of the response phase.

Latency

Correlations between CDI and latency of first look to the target are in Table 4 (Measure 2). As expected, correlations for both noun trials and verb trials are negative, indicating that toddlers with higher vocabularies identify the target more quickly. However, the correlation is only statistically significant for nouns.

---

[4]The large range and standard deviation for nouns are likely due to the small number of noun trials in comparison to the number of verb trials.

Fixation Density

We correlated toddlers' CDI score with fixation density or the proportion of the number of fixations within the distractor scene to the number of fixations within the target scene in both the short and long time windows. We found no reliable results for the long window, but we found reliable results for both nouns and verbs in the short window (see Table 4 [Measure 3]). This is sensible given that, over the course of the trial, toddlers' gaze patterns stabilized; only at the beginning, during the short time window, were the effects of "checking back" likely to be evident.

For nouns, the correlation is negative, meaning that, as CDI score increases, the proportion of distractor to target fixations decreases, whereas the opposite holds for verbs. This suggests the following. On noun trials, toddlers who had larger expressive vocabularies made several individual fixations on the target and made few looks to the distractor image, there being no need to check back once toddlers confirmed that the static distractor image was incorrect. On verb trials, toddlers with larger expressive vocabularies continually inspected the distractor scene, perhaps to check that it was in fact not the named target or to understand how it differed from the other event (e.g., Fitneva & Christiansen, 2011). This does not mean that these toddlers preferred the target overall. On the contrary, the correlation between increased looking time from baseline to test and CDI for verb trials was positive, and significantly so in the 6-s time window, indicating that they preferred the target overall. Instead, the proportion reveals that they looked around the distractor scene a greater number of times than they did on the target scene, but with short durations—these looks were brief "checking" looks. For both nouns and verbs, toddlers with smaller vocabularies, that is, those who likely knew fewer of the queried words, were more likely to look back and forth at both scenes at a similar rate.

Discussion

In the past 25 years or so, research on verb acquisition has exploded, fueling new theoretical developments and empirical advances (see, e.g., Hirsh-Pasek & Golinkoff, 2006; Syrett & Arunachalam, 2018; Tomasello & Merriman, 1995). Nevertheless, the majority of the field's understanding of early vocabulary knowledge comes from research about nouns. The current study was an eye-tracking investigation of receptive verb knowledge using dynamic video scenes to depict verbs. We tested a relatively large number of verbs and focused on a key but understudied age range in which toddlers are actively building their verb vocabularies. We also included noun trials using static images to provide a link to prior research (e.g., Fernald et al., 1998; Houston-Price et al., 2007) and to be able to address how children respond in an eye-gaze task to static images depicting nouns versus dynamic scenes depicting verbs. Characterizing the data broadly, we found that, overall, toddlers as a group knew just over half of the tested verbs (by one commonly used measure—a 15% increase in looking time

Figure 5. Toddlers' preference for the target scene on verb trials, grouped by the scene pairings on each trial and the phase of the trial: baseline versus response. Error bars indicate standard error of subject means.
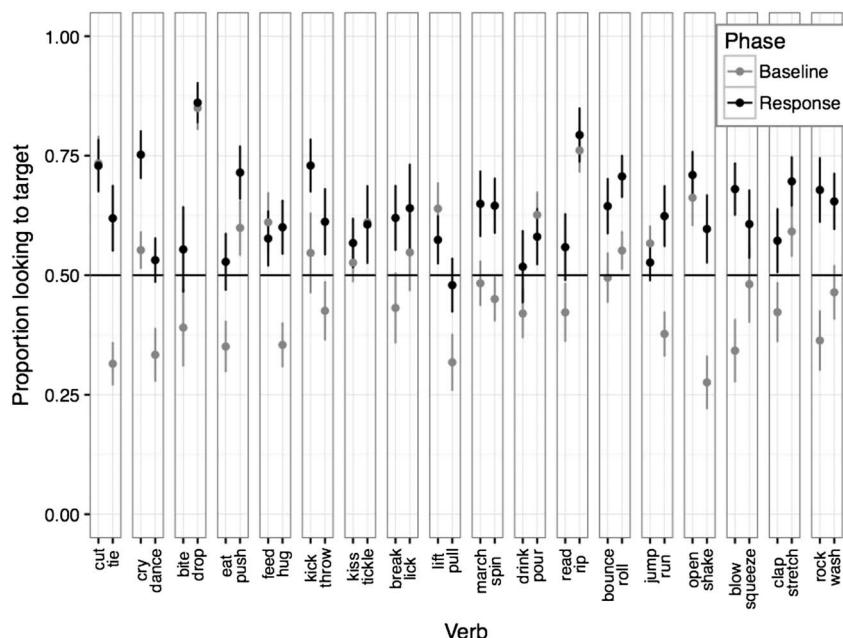
to the target after the target verb is provided—within a time window of 300–6000 ms of the test period). We used stimuli from Konishi et al. (2016), who designed them for a pointing task with 30-month-olds and found that children knew 80% of the 36 tested verbs. We reported in Table 1 that parent report of expressive verb vocabulary was, on average, only five of the 14 tested verbs. Thus, the finding that toddlers at 22–24 months of age have 55% of the 18 verbs they were tested on in their receptive vocabularies is reasonable, and it is consistent with prior evidence that receptive vocabulary outstrips expressive vocabulary from as early as 6 months of age (Bergelson & Swingley, 2012). Thus, although we use eye tracking instead of pointing, the results are unsurprising and suggest that toddlers' vocabularies grow rapidly between the handful of verbs they appear to know in infancy (Bergelson & Swingley, 2013) and the (at least) 18 verbs they are reported to know in Konishi et al.'s work.

The primary goal of this work was to understand methodological aspects of assessing individual toddlers' receptive verb knowledge using eye-gaze. We suspected that, because dynamic scenes are likely to capture visual attention in different ways from static images, the standard eye-gaze measures typically used to study static images in prior work might not work in the same way for dynamic scenes. This is an important methodological question as the field moves forward. Should these measures prove differentially effective for measuring noun and verb knowledge, it is crucial that researchers use the best measures for their purpose. Therefore, we explored three ways of examining

Table 4. Correlations between gaze measures and score on the MacArthur-Bates Communicative Development Inventories: Words and Sentences, Short Form A (a measure of expressive vocabulary), for nouns and verbs.

	1				2		3			
	Proportion of trials with ≥ 15% target preference increase				Latency of the first look to the target (excluding the first 50 ms)		Fixation density			
	300–1800 ms		300–6000 ms				300–1800 ms		300–6000 ms	
Measure	r	p	r	p	r	p	r	p	r	p
Nouns	.36	.02*	.26	.1	−.33	.04*	−.36	.03*	−.20	.2
Verbs	.12	.2	.37	.02*	−.27	.09	.35	.03*	−.13	.4

Note. Measures 1 and 3 are each evaluated in two time windows.

*p < .05.

the gaze data, correlating each gaze measure with toddlers' expressive vocabulary on the CDI.

The results did indeed indicate that there are subtle differences in how static noun and dynamic verb measures should be operationalized. For gauging accuracy, or increased looking time from baseline to test, although we found the expected correlations, the duration of the analysis window mattered. For nouns, we saw a correlation with CDI in a short time window from 300 to 1800 ms of the test window (a common window in such studies, e.g., Fernald et al., 2008) but not the full 300- to 6000-ms window. For verbs, we found the opposite pattern: a reliable correlation obtained in the full window but not the shorter window. These results are consistent with prior work showing that an increase in preference for the target is a useful measure of both noun and verb knowledge (Goldfield et al., 2016; Killing & Bishop, 2008; Reznick, 1990; Reznick & Goldfield, 1992) but further indicate that, for toddlers under 2 years of age, dynamic verb trials necessitate a longer analysis window than static noun trials. We think this is a reflection of the dynamicity of the scenes on verb trials—on noun trials, toddlers could look immediately at the target static image and then look away when they were bored of it, whereas on verb trials, the motion in both scenes captured their attention throughout, resulting in a slow accumulation of increased looking to the target over the distractor.

For latency, we found that only noun performance correlated with CDI; verb performance showed no correlation. Thus, for the static noun trials, we replicate the pattern in prior work by Fernald and colleagues (e.g., Fernald et al., 2001, 2008; Marchman & Fernald, 2008), but for the dynamic verb trials, latency may be less useful for assessing verb knowledge in an eye-tracking task, at least for this age group. We suggest this is because the first fixation to dynamic scenes, unlike static objects, may be determined more by movement-related features rather than lexical knowledge, at least for toddlers at this age.

We also pursued a third measure, borrowed from the literature on complex scene viewing in adults, but not, to our knowledge, used to study child language: fixation density or the proportion of the number of fixations to the distractor as compared to the number of fixations to the target (e.g., Henderson et al., 1999). This measure correlated with vocabulary score on the CDI for both static noun and dynamic verb trials, albeit in opposite directions. For noun trials, there was a reliable negative correlation between vocabulary score on the CDI and the proportion of fixations to the distractor relative to the target, whereas for verb trials, there was a reliable positive correlation. This finding suggests, first, that fixation density is a useful measure for assessing individual differences in vocabulary knowledge and, second, that dynamic scenes depicting verbs and static images depicting nouns show entirely different patterns of attention: Static images depicting nouns lead toddlers to launch fixations within the target region at a high rate, whereas dynamic scenes depicting verbs lead them to launch (brief) fixations within the distractor region

at a high rate. We suggest that this is because toddlers needed to continually verify that the distractor scene would not change to depict the target event. Notably, the correlations between fixation density and vocabulary score for both nouns and verbs only held for the first 1500 ms of the time window, suggesting that fixation patterns evened out over the course of each trial. The fixation density results are intriguing but require replication with larger samples across different ages before the potential utility of this measure can be fully understood.

Limitations

Two limitations of this study that we hope to address in future work involve the stimuli. First, the fact that we depicted nouns with static images and verbs with dynamic scenes means that we cannot tell whether differences in performance on noun and verb trials were due to differential knowledge of nouns and verbs or to differences in gaze to static images versus dynamic scenes. We used static images for noun trials to provide a link to the existing literature, and of course, we used dynamic scenes for verb trials because this is the best way to depict verb referents (Golinkoff et al., 1987). In future work, it would be useful to target noun referents with dynamic scenes in which actors engage with the targeted object referents. Including similarly rich and complex scenes for both noun and verb referents would reveal how dynamicity itself, and embedding in an action scene, affects gaze behavior.

Second, although preference for the target scene during the baseline phase, before the test query, was 50% overall, reflecting equal overall preference for the target and the distractor, Figure 5 indicates that some scenes were much more salient than the others, perhaps even bringing toddlers to ceiling levels, which affected our ability to detect knowledge of the verb meanings during the response phase. Because the stimuli were originally designed for a pointing task with older children, for which salience may matter less given that they are providing an overt response, it may be appropriate to edit some of the video stimuli for future eye-tracking studies to ensure that they are more equal in salience.

A limitation of the procedure is that we did not include other measures of toddlers' developmental level. We used the CDI as the comparison point for judging toddlers' vocabulary knowledge, and the correlations with eye-gaze measures for nouns, as well as for both nouns and verbs on fixation density, suggest that there is a relationship between these measures. We noted that some kind of comparison point is required for this study and that it is possible that, to the extent that parent report is suboptimal, it may be so across the board. Nevertheless, it is also possible that caregivers differ in their ability to correctly estimate toddlers' expressive vocabularies, making this a limitation of our approach. Further, given our focus on verbs, it would be particularly useful to see how the gaze measures relate to syntactic knowledge (of which the only measure was the question on the CDI about whether toddlers are combining

words—and all were, either "sometimes" or "often"). Future work should use a battery of assessments to explore how verb vocabulary relates to other linguistic and cognitive abilities in the toddler years.

One potential limitation, but also perhaps an advantage, of the current study's approach is that eye tracking permits study of emergent vocabulary knowledge. Both Houston-Price et al. (2007) and Robinson, Shore, Hull Smith, and Martinelli (2000) found that toddlers in IPLP tasks looked at noun targets even when their parents indicated their toddlers did not know those nouns. Of course, this could be because parents underestimate their toddlers' knowledge. However, many researchers have suggested that IPLP tasks can tap into representations that are not semantically rich or detailed (e.g., Chang, Dell, & Bock, 2006; Chita-Tegmark, Arunachalam, Nelson, & Tager-Flusberg, 2015). Thus, the current task may not reveal whether toddlers have deep or sophisticated knowledge of the verbs' meanings but rather whether toddlers identify the target as a better referent than the distractor scene. We think this is an advantage: Understanding emergent knowledge is likely to offer early predictive information about individual toddlers' language development. Nevertheless, with the current design, we cannot explore important questions about toddlers' abilities to extend verbs to new contexts and new uses (e.g., Naigles et al., 2009), which have been the target of much existing work with both familiar verbs (e.g., Forbes & Poulin-Dubois, 1997; Seston, Golinkoff, Ma, & Hirsh-Pasek, 2009) and novel verbs (e.g., Arunachalam & Waxman, 2015; Poulin-Dubois & Forbes, 2002; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). Indeed, it may be that toddlers' knowledge is underestimated when they are required to identify referents that have different actors, settings/backgrounds, or objects than they are used to (Naigles & Hoff, 2006), but this would be a problem for virtually any kind of standardized assessment other than parent report.

Conclusion

The study's results indicate, first, that typically developing toddlers aged 22–24 months have robust but still growing verb lexicons. Second, the results suggest that an eye-tracking test for verb vocabulary is within reach. We believe we have begun the process of refining the operationalization of gaze measures that could be used in such a test, but much work still needs to be done. Such a test, once developed, would have an important clinical impact. Eye tracking has been lauded as an excellent technique for assessing receptive vocabulary knowledge in clinical populations, including children with motor impairments such as cerebral palsy (Cauley, Golinkoff, Hirsh-Pasek, & Gordon, 1989) and children with social communication impairments such as autism spectrum disorder (e.g., Brady, Anderson, Hahn, Obermeier, & Kapa, 2014; Chita-Tegmark et al., 2015; Horvath & Arunachalam, 2018; Swensen, Kelley, Fein, & Naigles, 2007; Venker & Kover, 2015). Eye-tracking tasks may be better than either parent report or child-based assessments for children with these developmental disabilities

who may, for example, struggle with the behavioral task demands of the PPVT and whose receptive knowledge may be vastly underestimated by the CDI (e.g., Rapin, Dunn, Allen, Stevens, & Fein, 2009; Tager-Flusberg & Kasari, 2013).

In particular, given that language and communication disorders are often associated with selective impairment of verb knowledge (e.g., Fletcher, 1994; Hadley, 2006; Menyuk & Quill, 1985; Olswang et al., 1997; Rice & Bode, 1993), it is important to have tools to accurately assess verb knowledge in children at risk for or diagnosed with disorders. In addition, given the importance of early intervention for optimal outcomes, verb assessments that are appropriate for young children, particularly at the ages when late talkers can be identified, are ideal. Eye tracking during dynamic scenes is appropriate for young children and provides ecologically valid depictions of verb meanings—and thus may have an important role to play in developing such assessments.

References

Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes, 28*(4), 417–425. https://doi.org/10.1080/01690965.2011.641744

Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development, 7*(1), 169–184. https://doi.org/10.1080/15475441.2011.573760

Arunachalam, S., & Waxman, S. R. (2015). Let's see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition, 22*(2), 117–131. https://doi.org/10.1080/10489223.2014.928300

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

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

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. https://doi.org/10.1073/pnas.1113380109

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

Bernal, S., Dehaene-Lambertz, G., Millotte, S., & Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Developmental Science, 13*(1), 69–76. https://doi.org/10.1111/j.1467-7687.2009.00865.x

Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye tracking as a measure of receptive vocabulary in children with autism spectrum disorders. *Augmentative and Alternative Communication, 30*(2), 147–159. https://doi.org/10.3109/07434618.2014.904923

Candan, A., Küntay, A. C., Yeh, Y., Cheung, H., Wagner, L., & Naigles, L. R. (2012). Language and age effects in children's processing of word order. *Cognitive Development, 27*(3), 205–221. https://doi.org/10.1016/j.cogdev.2011.12.001

Cauley, K. M., Golinkoff, R. M., Hirsh-Pasek, K., & Gordon, L. (1989). Revealing hidden competencies: A new method for studying language comprehension in children with motor impairments. *American Journal of Mental Retardation, 94*(1), 53–63.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chita-Tegmark, M., Arunachalam, S., Nelson, C. A., & Tager-Flusberg, H. (2015). Eye-tracking measurements of language processing: Developmental differences in children at high risk for ASD. *Journal of Autism and Developmental Disorders, 45*(10), 3327–3338. https://doi.org/10.1007/s10803-015-2495-5

Cocking, R. R., & McHale, S. (1981). A comparative study of the use of pictures and objects in assessing children's receptive and productive language. *Journal of Child Language, 8*(1), 1–13. https://doi.org/10.1017/S030500090000297X

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the Intermodal Preferential Looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior & Development, 40,* 151–172. https://doi.org/10.1016/j.infbeh.2015.05.005

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test (PPVT)*. Minneapolis, MN: Pearson Assessments.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 174–185. https://doi.org/10.2307/1166093

Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (pp. 97–123). Philadelphia, PA: John Benjamins.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by

infants in the 2nd year. *Psychological Science, 9*(3), 228–231. https://doi.org/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. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics* (pp. 97–135). Philadelphia, PA: John Benjamins.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology, 31*(1), 41–81. https://doi.org/10.1006/cogp.1996.0012

Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science, 5*(1), 55–64. https://doi.org/10.1111/1467-7687.00209

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science, 35*(2), 367–380. https://doi.org/10.1111/j.1551-6709.2010.01156.x

Fletcher, P. (1994). Grammar and language impairment: Clinical linguistics as applied linguistics. In D. Graddol & J. Swann (Eds.), *Evaluating language* (pp. 1–21). Clevedon, United Kingdom: Multilingual Matters/BAAL.

Forbes, J. N., & Poulin-Dubois, D. (1997). Representational change in young children's understanding of familiar verb meaning. *Journal of Child Language, 24*(2), 389–406.

Friedman, S. L., & Stevenson, M. B. (1975). Developmental changes in the understanding of implied motion in two-dimensional pictures. *Child Development, 46*(3), 773–778. https://doi.org/10.2307/1128578

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, 35*(2), 302–309. https://doi.org/10.3758/BF03202556

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). New York, NY: Oxford University Press.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*(8), 684–691. https://doi.org/10.1111/j.1467-9280.2006.01767.x

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goldfield, B. A., Gencarella, C., & Fornari, K. (2016). Understanding and assessing word comprehension. *Applied Psycholinguistics, 37*(3), 1–21. https://doi.org/10.1017/S0142716415000107

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language, 17*(1), 171–183. https://doi.org/10.1017/S0305000900013167

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences, 12*(10), 397–403. https://doi.org/10.1016/j.tics.2008.07.003

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45. https://doi.org/10.1017/S030500090001271X

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the Intermodal Preferential Looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science, 8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Hadley, P. A. (2006). Assessing the emergence of grammar in toddlers at risk for specific language impairment. *Seminars in Speech and Language, 27*(3), 173–186. https://doi.org/10.1055/s-2006-948228

Hadley, P. A., Rispoli, M., & Hsu, N. (2016). Toddlers' verb lexicon diversity and grammatical outcomes. *Language, Speech, and Hearing Services in Schools, 47,* 44–58. https://doi.org/10.1044/2015_LSHSS-15-0018

He, A. X., & Arunachalam, S. (2017). Word learning mechanisms. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(4), e1435. https://doi.org/10.1002/wcs.1435

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The preferential looking paradigm reveals emerging language comprehension. In D. McDaniel, C. McKee, & H. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). Cambridge, MA: MIT Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (2006). *Action meets word: How children learn verbs.* Oxford, United Kingdom: Oxford University Press.

Horvath, S., & Arunachalam, S. (2018). *Individual differences in eye-gaze measures of receptive vocabulary*. Manuscript in preparation.

Horvath, S., Rescorla, L., & Arunachalam, S. (2018). Semantic features of early verb vocabularies. In K. Syrett & S. Arunachalam (Eds.), *Semantics in Language Acquisition* (pp. 67–93). Amsterdam, Netherlands: John Benjamins.

Horvath, S., Rescorla, L., & Arunachalam, S. (In press). The syntactic and semantic features of two-year-olds' verb vocabularies: A comparison of typically developing children and late talkers. *Journal of Child Language*.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language, 34*(4), 701–724. https://doi.org/10.1017/S0305000907008124

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, *90*(1), 72–93. https://doi.org/10.1037/0033-295X.90.1.72

Killing, S. E. A., & Bishop, D. V. M. (2008). Move it! Visual feedback enhances validity of preferential looking as a measure of individual differences in vocabulary in toddlers. *Developmental Science, 11*(4), 525–530. https://doi.org/10.1111/j.1467-7687.2008.00698.x

Konishi, H., Stahl, A. E., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Individual differences in nonlinguistic event categorization predict later motion verb comprehension. *Journal of Experimental Child Psychology, 151,* 18–32. Advance online publication. https://doi.org/10.1016/j.jecp.2016.03.012

Li, P. (2001). *The CHILDES Parental Corpus*. Retrieved from https://childes.talkbank.org/derived/parentfreq.cdc

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11*(4), 628–634. https://doi.org/10.1111/j.1467-7687.2008.00707.x

Marchman, V. A., & Bates, E. A. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339–366. https://doi.org/10.1017/S0305000900009302

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), 9–17. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three- and five-year-old children. *Journal of Child Language, 35*(2), 373–402. https://doi.org/10.1017/S0305000907008549

Menyuk, P., & Quill, K. (1985). Semantic problems in autistic children. In E. Schopler & G. Mesibov (Eds.), *Communication problems in autism* (pp. 127–145). New York, NY: Plenum Press.

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development, 11,* 356–368. https://doi.org/10.1080/15475441.2014.978331

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*(2), 357–374.

Naigles, L. (1997). Are English-speaking one-year-olds verb learners, too? In E. Clark (Ed.), Proceedings of the 28th Annual Child Language Research Forum (pp. 199–212. Stanford, CA: CSLI.

Naigles, L. R., & Hoff, E. (2006). Verbs at the very beginning: Parallels between comprehension and input. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 336–365. New York, NY: Oxford University Press.

Naigles, L. R., Hoff, E., & Vear, D. (2009). Flexibility in early verb use: Evidence from a multiple-N diary study. *Monographs of the Society for Research in Child Development, 74*(2), 1–143. https://doi.org/10.1111/j.1540-5834.2009.00513.x

Olswang, L., Long, S., & Fletcher, P. (1997). Verbs in the emergence of word combinations in young children with specific expressive language impairment. *European Journal of Disorders of Communication, 32,* 15–33. https://doi.org/10.1111/j.1460-6984.1997.tb01622.x

Poulin-Dubois, D., & Forbes, J. N. (2002). Toddlers' attention to intentions-in-action in learning novel action words. *Developmental Psychology, 38*(1), 104–114. http://doi.org/10.1037/0012-1649.38.1.104

Rapin, I., Dunn, M., Allen, D., Stevens, M., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology, 34,* 66–84.

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rescorla, L. A. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54,* 587–599. https://doi.org/10.1044/jshd.5404.587

Rescorla, L. A., & Dale, P. S. (2013). *Late talkers: Language development, interventions, and outcomes*. Baltimore, MD: Brookes.

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics, 11*(2), 145–166. https://doi.org/10.1017/S0142716400008742

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology, 28*(3), 406–413. https://doi.org/10.1037/0012-1649.28.3.406

Rice, M., & Bode, J. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language, 13*(37), 113–131. https://doi.org/10.1177/014272379301303707

Robinson, C. W., Shore, W. J., Hull Smith, P., & Martinelli, L. (2000, July). *Developmental differences in language comprehension: What 22-month-olds know when their parents are not sure.* Poster session presented at the International Conference on Infant Studies, Brighton, England.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium, 71–78.* https://doi.org/10.1145/355017.355028

Seston, R., Golinkoff, R. M., Ma, W., & Hirsh-Pasek, K. (2009). Vacuuming with my mouth?: Children's ability to comprehend novel extensions of familiar verbs. *Cognitive Development, 24*(2), 113–124. https://doi.org/10.1016/j.cogdev.2008.12.001

Styles, S., & Plunkett, K. (2009). What is "word understanding" for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language, 36*(4), 895–908. https://doi.org/10.1017/S0305000908009264

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: Evidence from preferential looking. *Child Development, 78*(2), 542–557. https://doi.org/10.1111/j.1467-8624.2007.01022.x

Syrett, K., & Arunachalam, S. (2018). *Semantics in acquisition* (trends in language acquisition research). Amsterdam: John Benjamins.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6), 468–478.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development, 70*(3), 620–635. https://doi.org/10.1111/1467-8624.00045

Thordardottir, E. T., & Weismer, S. E. (1983). Verb argument structure weakness in specific language impairment in relation to age and utterance length. *Clinical Linguistics & Phonetics, 16*(4), 233–250. https://doi.org/10.1080/02699200110116462

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175. https://doi.org/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. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tomasello, M., & Merriman, W. (Eds.). (1995). *Beyond names for things: Young children's acquisition of verbs.* Hillsdale, NJ: Erlbaum.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development, 59*(5), 174–179.

Trueswell, J. C., & Gleitman, L. R. (2007). Learning to parse and its implications for language acquisition. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 635–656). Oxford, United Kingdom: Oxford University Press.

Venker, C. E., & Kover, S. T. (2015). An open conversation on using eye-gaze methods in studies of neurodevelopmental disorders. *Journal of Speech, Language, and Hearing Research, 58*(6), 1719–1732. https://doi.org/10.1044/2015_JSLHR-L-14-0304

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x