Title
Connecting the tots: Strong looking-pointing correlations in preschoolers' word learning and implications for continuity in language development

Author
Creel, Sarah C

Publication Date
2024-08-21

DOI
10.1111/cdev.14157

DOI: 10.1111/cdev.14157

CHILD DEVELOPMENT

EMPIRICAL ARTICLE

Connecting the tots: Strong looking-pointing correlations in preschoolers' word learning and implications for continuity in language development

Sarah C. Creel

Department of Cognitive Science, La Jolla, California, USA

Correspondence
Sarah C. Creel, Department of Cognitive Science, 9500 Gilman Drive Mail Code 0515, La Jolla, CA 92093-0515, USA.
Email: screel@ucsd.edu

Funding information
Division of Behavioral and Cognitive Sciences, Grant/Award Number: BCS-1057080 and BCS-1230003

Abstract
How does one assess developmental change when the measures themselves change with development? Most developmental studies of word learning use *either* looking (infants) or pointing (preschoolers and older). With little empirical evidence of the relationship between the two measures, developmental change is difficult to assess. This paper analyzes 914 pointing, looking children (451 female, varied ethnicities, 2.5–6.5 years, dates: 2009–2019) in 36 word- or sound-learning experiments with two-alternative test trials. Looking proportions and pointing accuracy correlated strongly ($r=.7$). Counter to the "looks first" hypothesis, looks were not sensitive to incipient knowledge that pointing missed: when pointing is at chance, looking proportions are also. Results suggest one possible path forward for assessing continuous developmental change. Methodological best practices are discussed.

In many areas of developmental science, research procedures vary substantially between infancy and early childhood. This partly reflects the characteristics of participants at different ages: for example, 6-month-olds cannot provide interpretable verbal or pointing responses, while 3-year-olds are likely too mobile to tolerate a habituation procedure. Thus, infants are typically tested using looking measures: commonly, head turn duration to a sound source, visual habituation-dishabituation, and/or fixation duration to a named picture or surprising event. By contrast, children starting around age 3 years typically provide pointing, reaching, touching, or verbal responses. These apparently more overt responses are simpler to obtain than the more ambiguous visuomotor measures obtained with infants, which require either specialized equipment or laborious hand-coding of fixation locations and times.

Age-specialized approaches allow probing of developing knowledge within an age group. However, age-specialized approaches are less informative about developmental change, limiting our ability as a field to understand continuity or discontinuity in development (see Cowan, 2016; Creel & Quam, 2015; Keen, 2003; among others, on disconnects across age groups).

One way of bridging this developmental gap in understanding between disparate measures is to compare the two types of measures within subjects. My goal in this study was to make a first step in connecting looking and pointing measures of comprehension by examining a large amount of within-subjects looking and pointing data from children aged 2.5–6.5 years who are attempting to recognize the visual referents of newly learned words or other audiovisual associations. While the goal was to connect to infant research, notably, this type of investigation would be impossible with young infant data as infants cannot produce pointing responses reliably.

Efforts to understand language development offer a good case study of the dilemma of age-specific research procedures. For language researchers, a crucial assumption is that people spontaneously look at things that are being talked about (Cooper, 1974; Tanenhaus et al., 1995), and therefore looking at a thing that is being talked about indicates comprehension. Developmental language researchers have used

Abbreviations: BF, Bayes factor; RT, reaction time.

this to great advantage, demonstrating comprehension that greatly precedes production of language. Some research findings suggest word recognition at 6 months (Bergelson & Swingley, 2012, 2014; Tincoff & Jusczyk, 1999, 2012); yet others suggest difficulty in word learning at 3–5 years (e.g., Creel, 2014a; Creel & Frye, 2024; Storkel, 2001, 2003).

Studies of each age group often use nearly parallel research designs: During a learning phase, an object is labeled repeatedly; in a test phase, two pictures are presented at a time while one is named. However, the two age ranges use different dependent measures: looking for infants, and pointing for young children. Somewhat surprisingly, *we do not know how closely looking time and pointing measures of comprehension correspond to each other*, even though the research designs are similar at an abstract level. This makes it difficult to assess gains, plateaus, or losses during development. To illustrate concretely, suppose 6- to 9-month-old infants hearing *ball* look at the ball picture on average 52% of the time and the other picture 48% of the time in a 4- or 5-s-long time window, whereas 3-year-olds point to the ball 90% of the time (52% looks estimate derived from time course plots in Bergelson & Aslin, 2017; Bergelson & Swingley, 2014, using WebPlotDigitizer 4.2; Rohatgi, 2019). One can say that the 6- to 9-month-olds are statistically above chance and therefore know what a ball is, and that the 3-year-olds are statistically above chance and therefore know what a ball is. However, this binary between "know" and "doesn't know" runs counter to theories of language development as a continuous process (e.g., Fernald et al., 2006; Hazan & Barrett, 2000; McMurray, 2007) and does nothing to inform the study of developmental change. That is, we currently cannot say whether these data indicate a developmental gain in word knowledge or not.

Complicating the interpretation of the relationship between looking and overt measures, one line of thinking is that looking measures may be more sensitive to partial or probabilistic associative knowledge than other, large-motor movements, especially in the youngest children, and especially when a task is difficult or unfamiliar. Looking is metabolically cheaper (McMurray & Aslin, 2004; eyes are smaller and less heavy than arms) and typically precedes and contributes to the execution of manual actions such as reaching or pointing (e.g., Droll & Hayhoe, 2007; Land et al., 1999). Supporting this "looks first" perspective, multiple studies in varied areas of developmental science report more advanced performance in looking measures compared to reaching, touching, or pointing measures. Vishton et al. (2005) reported that 6- to 7-month-olds retain object knowledge better in a looking measure than in a reaching measure. Cuevas and Bell (2010) found that infants aged 5–8 months showed more advanced performance in the looking version of the A-not-B task than the reaching version, whereas by 9–10 months, performance was comparable. Lee and Kuhlmeier (2014) found that 2-year-olds reliably looked to the correct falling location of a ball, but only some of them also reliably pointed to the correct location.

In the domain of word recognition specifically, Hendrickson et al. (2015, 2017; see also Smolak et al., 2021) reported that children who made a haptic (touchscreen) response in a two-alternative picture vocabulary test showed faster visual reaction times (RTs; defined as a move from the initially fixated incorrect picture to the correct picture) than when they made *no* haptic response. They inferred a possible intermediate knowledge state that supports looking but is not strong enough to support touching. Relatedly, Gurteen et al. (2011) reported that 13- and 17-month-olds showed knowledge of recently learned words in looking responses but not pointing responses.

Not all results agree on a looking advantage. For example, Diamond (1995) found comparably good visual and reaching performance in infants of matched ages in carefully equated tasks, and Frank et al. (2016) found roughly equivalent performance on eye tracking and tablet-based tasks in separate groups of 1- to 4-year-olds. In a word-learning paradigm with 17- to 31-month-olds, Bakopoulou et al. (2023) tested both looking and selection in the same children and found a strong relationship between children's likelihood of *looking at* a shape match to a named object and their probability of *physically selecting* that shape match. Hayne et al. (2016) presented two groups of children with familiar/unfamiliar picture pairs, such as one's own dog versus someone else's dog. For one group, they gave no verbal instruction, simply measuring familiar picture looking preference. For the other group, they instructed the child to (e.g.) "point to your dog" and measured familiar picture *pointing* proportion. Hayne et al. (2016) reported that 3-year-old passive lookers showed null familiarity preferences for some categories, yet 3-year-old pointers showed high pointing accuracy to familiar pictures. That is, pointing was "better" (more indicative of knowledge) than looking. Still, few of the described studies other than Bakopoulou et al. (2023) and Hendrickson et al. (2015, 2017; see Berthier et al., 2001; Von Hofsten et al., 1998, for more exceptions) obtained both looking and manual measures from the same children within the same task.

The current work

While most existing data on word learning in children past infancy use overt behavioral responses such as pointing, some studies in recent decades, including work in my lab, have collected both looking and pointing measures in young children aged 2.5–6.5 years. This is akin to widespread collection of both accuracy and reaction time (RT) measures in adult studies. Directly comparing looks and points within-subjects across many studies could shed light on the relationship between looking and pointing as dependent measures. This investigation compiles these data to yield insights on the (development of the) relationship between looking and pointing measures of recognition, which speaks to the "meaning" of infant looking.

Again, my goal was to facilitate interpretive comparison of looking and pointing responses to words by infants and children by compiling and examining within-subjects pointing and looking data from young children tested by my lab over the past decade and a half. I ask two major theoretical questions (below), and report a few exploratory findings that should be of interest to researchers. I consider the two major questions to be predominantly, but not completely, confirmatory; predicted outcomes are clear and simple. Exploratory measures are, as stated, exploratory: the large dataset afforded unforeseen insights into additional questions.

> *Question 1: What is the strength and nature of the relationship between pointing and looking measures in response to an external signal (e.g., a word)?* A weak relationship would suggest that trying to connect research on looking measures to research on overt recognition measures like pointing is problematic. A strong relationship, however, would suggest that these two dependent measures might be connected sensibly.
> *Question 2: Do looks detect associative knowledge that pointing does not?* That is, if a child displays chance pointing accuracy, do they nonetheless look at the correct referent more often? If so, this would support a view that looking measures tap an earlier, perhaps implicit level of knowledge that pointing cannot access. If not, this would support a view that looking measures reflect something like overt knowledge in young children.

Studies presented here use newly learned associations, such as novel word–picture associations. The primary interest is studies of word learning, but I draw additional data from analogous studies of nonspeech audiovisual association learning. While it is also of interest to test responses to familiar words, they are omitted from the central analyses for three reasons. First, using familiar words necessarily gives older participants an advantage: a 6-year-old has heard the word "fish" about twice as often as a 3-year-old has. Thus, testing all participants on newly learned associations better matches frequency of exposure across age. Second, using familiar words results in a large proportion of children at ceiling for pointing accuracy, which restricts the range of variation of that measure. Third, most of the familiar-word data my lab has collected presents four pictured alternatives rather than two, making eye gaze patterns somewhat less comparable to the two-alternative learning trials. Finally, there is a reasonable argument that newly learned words for non-infants are the closest analog to infant "familiar" word recognition. That is, infants may still be learning words that we as adults classify as highly familiar to children. Nonetheless, at the encouragement of anonymous reviewers, I also examined looking-pointing correlations in a small set of two-alternative familiar-word trials that were interspersed with some of the word-learning trials in the reported studies (see Table S2; Figure S1).

Each included study tested children's knowledge of newly learned audiovisual (e.g., word-picture) associations by presenting two pictures at a time, playing the sound, and obtaining both *proportion of gaze* to pictures while the name of one of them was spoken ("Where's the deev?") and children's *overt pointing responses* to the pictures. With the exception of a single study focused on 3-year-olds (see Table S1), all studies assessed children in a preschool age range, from age 3 years to around age 6.0, with a few children slightly below or slightly above that range. Studies were designed to appeal to young children, but within that range, no adjustments in design or stimuli were made to accommodate particular ages. Thus, study conditions alone cannot explain any age differences in performance.

After a brief overview of the studies included, I attempt to answer each Question in turn, along with relevant exploratory findings. I then describe implications of the current findings for understanding increases in word knowledge during development and for conducting eye tracking research in children. Finally, I point out limitations and suggestions for future research.

METHOD

Thirty-six studies testing associative learning between words/voices/melodies and pictures were represented, with a total of 60 conditions (see Table S1) and 914 child participants, yielding about 1500 data points in experimental conditions.

Study inclusion and exclusion

I attempted to assess all eye tracking data from learning studies of young children by my laboratory since it was established (Creel, 2014a, 2014c, 2016; Creel & Frye, 2024; Creel & Jimenez, 2012; Frye & Creel, 2022), including some unpublished datasets (see list in Table S1). Some studies were excluded because I was uncertain of the coding of stimulus properties by graduate students who have since left the field. I further excluded from the main analysis studies or trials within studies where children looked and pointed to (potentially) familiar words, though see Table S2 and Figure S1. I also omitted studies where additional information presented earlier in a sentence hinted at the target prior to the word because asynchronous multiple cues might obscure the relationship between looking and pointing. For example, one set of studies

(Creel, 2014b) presented sentences with talker cues that potentially predicted object targets: "I want to see the marv" where only one of two talkers had previously labeled the *marv*.

Some studies included multiple pronunciations of words or conflicting cues to recognition (e.g., a melody associated with Object 1 but played on the musical instrument associated with Object 2). A study on affect recognition was included even though it had no learning component because children performed with similarly modest accuracy to many learning studies, suggesting that associations between affect and vocal properties are still being learned (see Friend & Bryant, 2000; Nelson & Russell, 2011; Quam & Swingley, 2012).

Participants

Data were collected in preschools and day cares in the greater San Diego, CA, US area that were contacted and were interested in research participation. Across studies (Table S1), 914 children were included (451 female, 455 male, 8 unstated; ethnicity varied but was not tracked; socioeconomic status was not tracked). When total *N*s are much larger, it is because *N* is the total number of participants *within each condition*; that is, participants who completed multiple conditions contributed multiple data points. Additional children were excluded if: their data were excluded from published studies; age data were missing; age <2.5 years or >6.5 years; eye tracking data did not meet criteria (see below). Given data recruitment patterns, it is possible that a few children took part in multiple studies over the multi-year period. However, this is relatively rare, would generally have been unrelated studies, and would have occurred at separations of 6+ months. Given the lengthy time elapsed since testing in most cases, such children were not tracked down and are not statistically treated as the same across studies.

Procedure

All studies were presented on a computer monitor (Figure 1). All but one study had one or more *learning phases* where a word (or other sound) was copresented multiple times with a visual associate. All studies had one or more *recognition phases* where a child heard a word or other auditory signal while viewing two pictures. As they heard the sound, their eye movements to pictures were tracked using an Eyelink 1000 eye tracker (SR Research, Mississauga, ON) in remote mode, sampling eye gaze position monocularly at 250 Hz or 500 Hz (2 or 4 ms samples). They were also asked to point to the named (or associated) picture, and an experimenter recorded their response via mouse click on the picture. In studies where a child completed multiple learn-recognize cycles, that child contributed multiple data points to the analysis and was treated as identical in following linear mixed effects models by including participant as a random intercept.

Sets of studies (see also Table S1)

Novel word learning (16 studies, $N = 545$)

In some studies, children completed learning of multiple word pairs in different learning-test sequences, such as first learning the words *deev* and *teev*, followed by a test on *deev* and *teev*; then they learned *vosh* and *vush*,

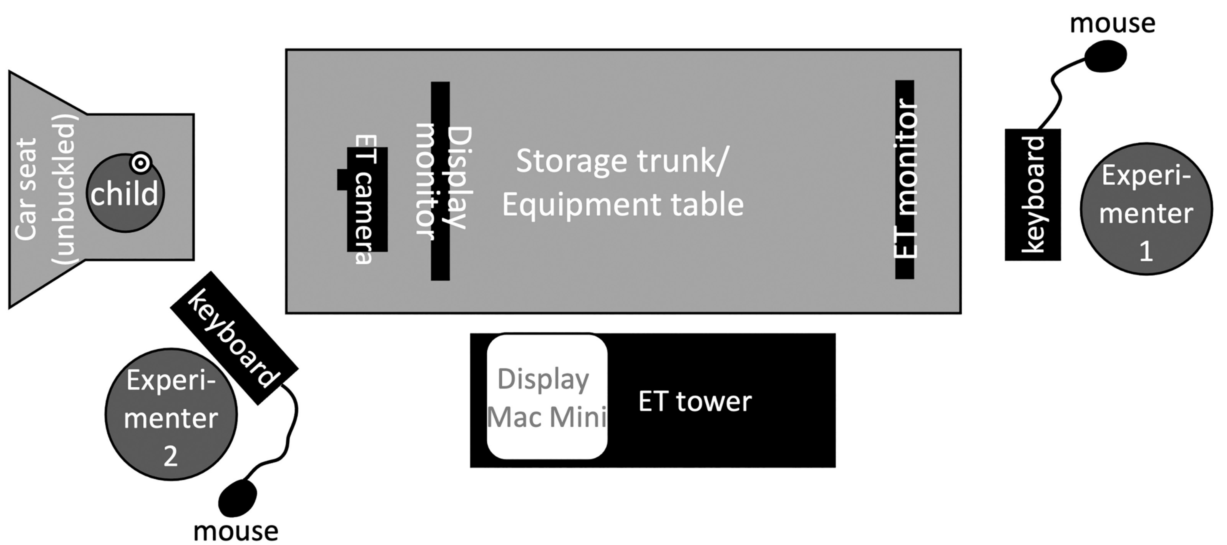

FIGURE 1 Experimental setup in preschool settings. The child sat in an unbuckled car seat (to maintain distance from eye tracker camera) facing the display monitor running MATLAB. The child wore a target-shaped sticker that allowed the Eyelink to determine head distance and position. The eye tracker camera was positioned below the display monitor. One experimenter monitored the eye tracking setup; the other sat next to the child to record their pointing responses via mouse click. ET, eye tracker.

followed by a *vosh*/*vush* test. In a case like this, the child contributed two data points, one for *deev*-*teev* (a consonant minimal pair) and another for *vosh*-*vush* (a vowel minimal pair).

Melody-picture association (13 studies, $N=239$)

Multiple studies assessed children's ability to form associations between visual shapes (cartoon characters) and brief melodies.

Talker-picture association studies (7 studies, $N=130$)

These examined children's ability to associate voices with pictures. One multi-experiment study was published (Creel & Jimenez, 2012; see also Jiménez & Creel, 2012) but omitted most of the eye tracking data. Further, not all talker learning participants had eye tracking data because at the time, all coauthors reasoned that pointing accuracy data were sufficiently revealing on their own and simpler to present. The other study was an honors thesis asking children to identify vocal affect (happy or sad) from isolated words.

Individual conditions within studies varied greatly in difficulty, with pointing accuracy ranging from .19 to .92 (Table S1). See Supporting Information for more detailed information on seemingly below-chance pointing.

Data preprocessing

Pointing accuracy and gaze data were reprocessed in a uniform way (described below), which in some cases differed from the previously published versions of the same studies. Details of data preprocessing appear in Supporting Information. Briefly, I wrote Python scripts to generate a database of each trial in the data set which grouped looks from word onset time (or other auditory event onset time) in 50-ms bins out to 3000 ms post-event onset. Using 50-ms bins rather than the original sampling rate (every 2–4 ms) greatly decreases the size of the data frame over which analyses are computed, thus accelerating computation speed. Averaged looks to images (target, i.e., correct answer; distractor) and other look categories were recorded, including a "bad looks" category (proportion off-screen looks or non-looks). Bins are plotted as ends of ranges; for example, the 50 ms bin included 1 through 50 ms.

Any blinks (really, any case where the tracker could not find the eye) were filled in with the most recent previous looking location, with filled-in status coded in a "blinks" column. This fill-in process was done because pointing sometimes obscures the eye image, yielding spurious not-looking values. Because fill-in was based on looking alone, it is logically independent of pointing and should not strengthen the relationship between looking and pointing. For any trial shorter than the maximum trial duration (set to 3000 ms), the final look location was extended to the maximum trial duration, as if the participant had kept looking until 3000 ms (though see later cautions on interpreting this very long time window). This prevents some trials from contributing more information than other trials at later time points.

After these preprocessing steps, I used R (R Core Team, 2020) for further processing and analyses of looking proportions and pointing accuracy. The looking measure was defined as proportion looks to target minus proportion looks to the other (distractor) picture, such that 0 = no bias to look toward either picture, during the time window from 200 to 2000 ms after target sound onset (see, e.g., Quam & Swingley, 2010, 2014, 2022; Swingley & Aslin, 2000, 2002; Yoshida et al., 2009, for use of very similar windows in infant looking time studies, as well as my work with 3- to 5-year-olds, Table S1; see Salverda et al., 2014, for evidence that the fastest signal-based fixations in adults are around 200 ms). Trials were excluded if 50% or more of looks were "bad," that is, not on the screen (including filled-in looks). After that, data cells were excluded if less than four trials were present after looking elimination. The four trial limit was chosen since some cells in some study designs had a maximum of only eight trials, thus 50% of trials retained. In principle one might vary this number to determine whether it has impacts on outcomes. The pointing accuracy proportion was the average pointing accuracy on trials that passed the looking exclusion criteria (rather than all trials before looking exclusions were applied), to provide the most straightforward test of the looking-pointing relationship.

Analysis

Question 1: What is the strength and nature of the relationship between pointing and looking measures in response to an external signal (e.g., a word)? I used linear mixed-effects regressions and correlations to examine relationships between each participant's accuracy and their looking proportions. Close correspondences between looking and pointing would suggest continuity across measures, and between looking proportion studies and overt pointing studies.

Question 2: Do looks detect associative knowledge that pointing does not? If looking is easier to plan and/or execute motorically than pointing, looks might access a more implicit level of knowledge. This hypothesis predicts that correct looks could be above chance even if pointing is at chance. This is assessed in two ways. First, I examine the intercept term in the earlier regression to gauge whether looks exceed chance where pointing is at chance. Second, I assess looking proportions of individuals whose pointing responses are at or around chance (i.e., 50%).

RESULTS

I first describe the overall outcome measures, including a check of the reliability of looking and pointing measures. I then turn to the specific research questions.

Reliability check

A preliminary question is how noisy the looking and pointing measures are. Following recommendations by Byers-Heinlein et al. (2022) for best practices in developmental research, I computed split-halves reliability of both looking (200–2000 ms window) and pointing accuracy via 5000 random splits of the data (Parsons et al., 2019) using the R package splithalf (version 0.8.2, Parsons, 2021). Split-halves reliability (Spearman-Brown corrected) for looking proportion was .50 (95% CI: [.45, .54]), and pointing accuracy was .63 (95% CI: [.60, .66]). Values were higher when limiting to the 35 conditions with ≥16 trials (looks: .59, 95% CI: [.45, .64]; points: .75, 95% CI: [.72, .78]), implying (unsurprisingly) that higher numbers of trials per participant increases reliability. Overall, these reliability values suggest that something repeatable is being measured.

General outcomes: Expected age-looking and age-pointing relationships, and strong looking-pointing relationship

Overall, both pointing and looking accuracy increased significantly with age, as expected (Table 1, upper; Figure 2, top and center; pointing × age raw correlation: $r(1493) = .150$, $p < .0001$; looking × age raw correlation: $r(1493) = .105$, $p < .0001$). Correlations were low due to substantial individual variability and cross-study variability in pointing and looking behavior, but age effects were highly robust in regression models (both $p < .0001$; see Supporting Information). Finally, and most importantly, looking and pointing correlated strongly *with each other* (Figure 2, bottom), $r(1493) = .709$, $p < .0001$. Correlations were weaker but qualitatively similar at the individual trial level (Table 1, lower). I also briefly examined familiar-word trials (217 data points from 125 children) and found a similarly strong correlation ($r = .642$), despite the fact that most data cells were at ceiling accuracy (62% 1.0 pointing accuracy; see Supporting Information for details). Finer grained analyses follow, but the large raw correlation between pointing and looking indicates a strong point-look relationship.

TABLE 1 Simple correlations.

	Averaged measures			Values on single trials		
	Looking proportion	**Pointing accuracy**	**Age (years)**	**Looking proportion**	**Pointing accuracy**	**Age (years)**
Looking prop.	.	.709	.105	.	.495	.048
Pointing acc.		.	.150		.	.079
Age (years)			.			.

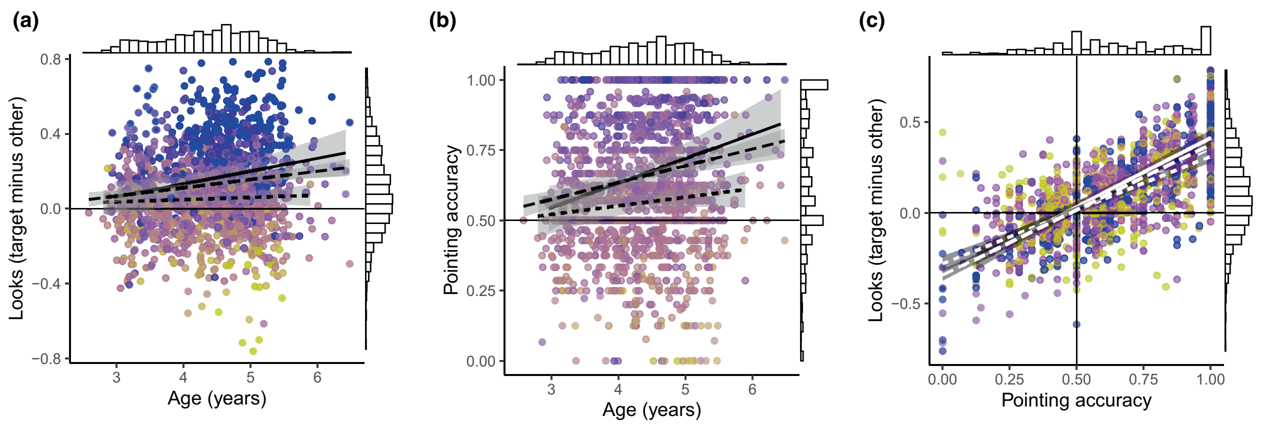

FIGURE 2 Relationships among age, pointing accuracy, and looking proportion (target minus other-picture looks, 200–2000 ms after auditory event onset; see text for details). Edge histograms show unidimensional distributions. Regression lines show 95% CIs; solid = talkers, dotted = melodies, and dashed = words. (a) Looks as a function of age. Pointing accuracy appears as a color gradient (low = lighter/yellow, high = darker/blue). (b) Pointing as a function of age, with looking proportion as a color gradient (low = lighter/yellow, high = darker/blue). (c) Relationship between pointing and looks. Yellow = age 3 years, purple = age 4 years, and blue = age 5 years.

Question 1: What is the strength and nature of the relationship between pointing and looking measures in response to an external signal (e.g., a word)? It is strong and positive

Correlations across individual conditions (Figure 3; Table S1) showed a median of .673 and a mean of .593. The mean, depressed by two >3.5 SD outliers, was .681 without outliers; 58/60 correlations were positive (binomial test $p < .0001$). These values were similar to the correlation across all data points (.709; Table 1). Thus, pointing accuracy explains up to 50.3% of the variance in looking accuracy. Further, within each experiment type (word learning, talker learning, and music learning) and most age groups, the pointing-looking relationship was strong (Figure S1). More accurate points mean more accurate looks.

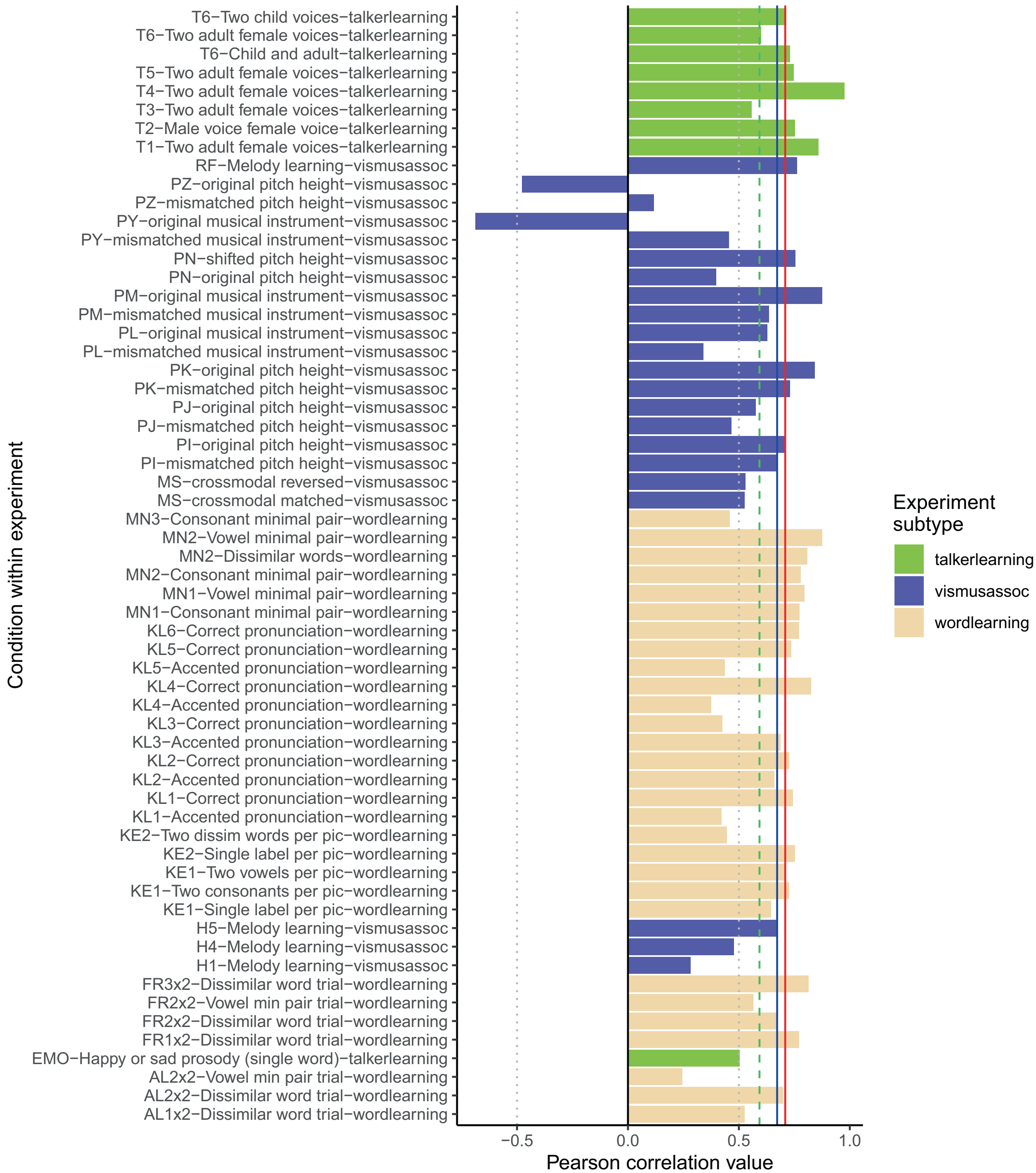

FIGURE 3 Looking-pointing correlations across all 60 experimental conditions, with brief descriptive titles. Vertical lines: Black = 0; dashed green = mean; blue = median; and red = value across all data points. Min = minimal; pic = picture.

Formalizing these observations, a linear mixed effects model used pointing accuracy to predict looks. Recall that the looking proportion measure has a true zero point, meaning equivalent looks to target and distractor. Pointing accuracy was recoded by subtracting .5, so that chance accuracy was 0. This means the intercept term will indicate whether looks exceed chance when accuracy is at chance (a point relevant for Question 2). Other factors included age (mean-centered); two scaled, centered variables contrasting the two non-word learning experiment subtypes with the word learning experiments; and random intercepts of participant (as some contributed multiple data points) and condition (i.e., a single condition within an experiment). Significance (alpha) values were calculated by likelihood ratio tests between the full model and otherwise identical models that held out the effect of interest. Of greatest interest were effects of pointing accuracy (indicating how much pointing accuracy predicts looking), age, and their interaction. Full model results appear in Supporting Information.

Pointing Accuracy significantly predicted looking proportion ($B=.656$, SE = .019, $t=34.64$, $p<.0001$), with more looks to the target as pointing accuracy increased. Age was not significant ($B=-.009$, SE = .005, $t=-1.69$, $p=.09$), but the Age × Pointing Accuracy interaction was ($B=.016$, SE = .019, $t=5.11$, $p<.0001$). A weak age effect may seem odd, but remember that this represents age differences *at the intercept* (chance pointing), rather than at higher pointing accuracy levels. The interaction speaks more directly to the question, do older children look more than younger children *matched for pointing accuracy*? To understand the interaction, separate models examined each year of age. As most of the sample ranged from 3.0 to 6.0, children under 3 years were classed with 3-year-olds and those over 6 years were classed with 5-year-olds. For 3-year-olds (2.58–3.99 years), Pointing Accuracy was significant ($B=.539$, SE = .037, $t=14.39$, $p<.0001$), showing increases in looks with increases in pointing accuracy, but the increase was smaller in magnitude than for 4-year-olds ($B=.682$, SE = .025, $t=27.15$, $p<.0001$), which was in turn smaller than for 5-year-olds (5.0–6.48 years; $B=.793$, SE = .038, $t=21.03$, $p<.0001$). This confirms age differences in looking-pointing slopes (see Figure S2). One reasonable guess is that younger children fixate targets more slowly than older children and are thus slower to arrive at "relevant" looks. However, an examination of how the looking-pointing correlation changes *over time* (Figure S3) suggests that younger groups' looks and points are less correlated throughout the duration of the trial. This implies that they may be less *consistent* lookers, exploring the display more than older children.

Exploration: What is the best time window to analyze?

Inherent in this question is whether the particular 200–2000 ms time window used here (and in much other developmental word recognition research) is either necessary or sufficient. Are other time windows more or less sensitive, and is the entirety of this range necessary? I addressed this by examining looking-pointing correlations in a growing window out to 3000 ms, as well as at four moving (sliding) window lengths (100 ms windows, 200 ms, 400 ms, and 800 ms) moved across the 200–3000 range in 100-ms increments (Figure 4).

Longer and later windows are more closely related to pointing responses

Looking-pointing correlations increase later in the trial, both with increasing windows and with sliding windows.

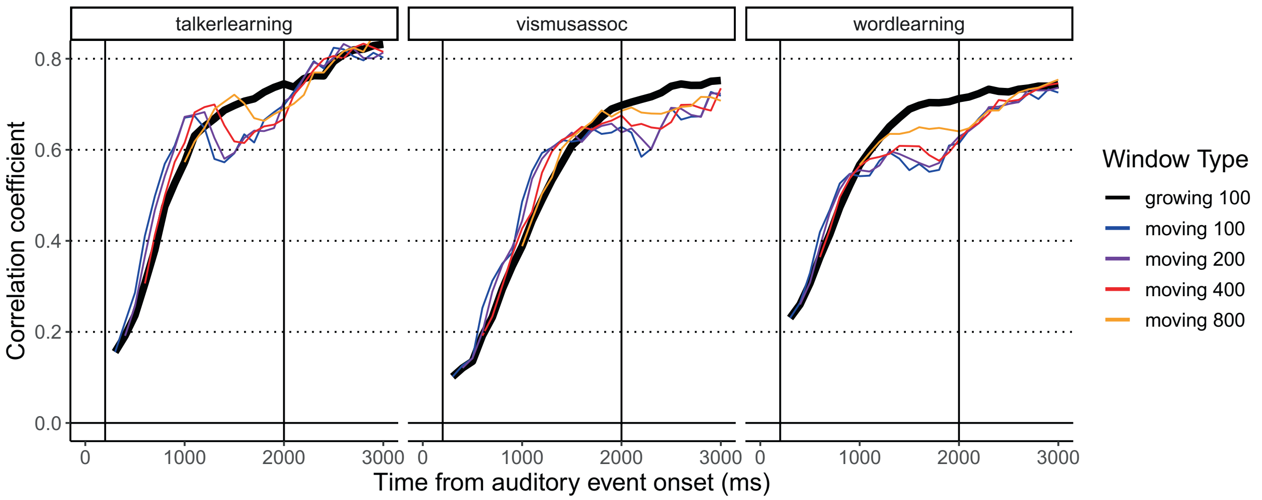

FIGURE 4 Correlations between accuracy and looks with increasingly long time windows (black lines) and moving windows of varying durations (100, 200, 400, and 800). The location of each point indicates the *end* of each time bin. For example, the first point on the yellow (moving 800 ms window) line encompasses 201–1000 ms and appears at 1000 on the *x* axis.

The relationship continues to increase as late as 3000 ms. However, the gain from 2000 ms to 3000 ms is limited. For moving windows, an initial steep rise is followed by a bumpy plateau that undershoots the growing window, then recovers. This general pattern is similar across the three different experiment types. If one wants to use looks as a proxy for overt recognition behavior, a long window may be a good choice; see Implications for Methodology below for further discussion.

Exploration: What do looking RTs mean? Perhaps not the same thing as pointing

Numerous studies (e.g., Fernald et al., 1998, 2001; Hendrickson et al., 2015, 2017; Marchman et al., 2022) have used looking RTs as a measure of word recognition speed. In this approach, the investigator examines only trials where the child initially fixated the non-target picture, and measures speed to shift away (presumably to the target). An advantage of looking RTs is that they are thought to measure rapidity of recognition. A disadvantage is that half or more of trials are excluded from analysis, potentially reducing power (though see Egger et al., 2020, for a gaze-contingent approach). It is not fully understood how looking RT relates to looking proportions or to pointing responses (though see Hendrickson et al., 2015, 2017).

I briefly examined looking RTs in the current data. To calculate it from the base dataset that contained each trial in consecutive 50-ms time bins, I first selected out distractor-initial trials and next, for each of those trials, calculated the time point at which the distractor was abandoned. I defined distractor-initial trials as those in which participants were fixating the nontarget picture for the entire 50-ms window right before sound-driven fixations might occur (151–200 ms). Looking RT was defined as the end of the first 50-ms bin with below-100% looks to the competitor, with the proportion of *non*-competitor looks × 50 ms subtracted from the bin end. For example, if the 50-ms bin ending in 350 ms contained 60% competitor looks, presumably the child moved off of the competitor 60% of the way into the bin, so $350 - (1 - 0.60) \times 50 = 330$ ms. About 38.6% of trials (7690 trials) contained a distractor-initial look (other locations included target, screen center, off-screen, and no looks). Of these, about 82.3% (6328 trials) shifted to the target before 2000 ms, similar to shift rates in Fernald et al. (2001) with much younger 21-month-old children. Trials that "timed out" (no target look before 2000 ms) were dropped from analysis.

Looking RTs related to pointing accuracy in two ways. First, trials with correct points were about 70 ms faster than trials with incorrect points (576 ms vs. 648 ms; Welch's $t(3209.1) = 7.36$, $p < .0001$). This confirms that faster look responses are more likely to reflect accurate pointing responses, consistent with use of this measure to index comprehension. Second, though, participants' overall looking RTs (limited to participants within a condition with four or more usable looking RTs) correlated only weakly (but significantly) with their averaged pointing accuracy on those trials (Table 2; $r(760) = -.140$, $p < .0001$). For maximum comparability, pointing accuracy values were limited to the trials used in the RT analysis.

The looking RT-pointing correlation of −.1 seems low compared to the overall correlation of pointing with looking *proportion* of around .7. Was this because a smaller number of trials were used in the current looking RT analysis versus the earlier looking proportion analysis? No. I calculated an average looking proportion measure limited only to this set of looking RT trials. Looking proportion still correlated with pointing accuracy more strongly than looking RTs did ($r(760) = .590$, $p < .0001$). Still, the two looking measures (RTs and proportion) correlated strongly ($r(760) = -.537$, $p < .0001$), suggesting that speed and overall looking proportion are related to each other. The relative weakness of the looking RT-pointing correlation may result from looks that happen *after* the looking RT on each trial; participants are more likely to look back away from the target on incorrect-point trials than on correct-point trials (see Figure S4 for a very cool trial-by-trial raster plot).

Does this mean that averaged looks are a more reliable measure of comprehension accuracy than looking RTs? This seems unlikely. Looking RT in infants correlates robustly during development with later language achievements (e.g., Fernald et al., 2006; Marchman & Fernald, 2008). One alternative interpretation is that studies here only contained a small number of RTs (5.45 ± 1.53 trials for word learning studies), while previous studies contained more (e.g., Fernald & Marchman, 2012: 19 trials per child). Fewer trials would result in higher RT noise here than in previous datasets. One such dataset in Peekbank (Zettersten et al., 2023), Adams et al. (2018), contained a large number of looking RT trials with familiar referents in 16- to 18-month-olds. I could not calculate relationship to pointing accuracy as children in that study did not point. However, a different way of assessing a measure's robustness is its reliability. In short, I retrieved Adams et al. (2018) from Peekbank using peekbankr (Braginsky et al., 2024) and extracted looking RTs. I found that reliability for looking RTs was .66, 95% CI: [.57, .73], while the comparable

TABLE 2 Correlations between looking reaction time (RT), looking proportion, and pointing accuracy on the same sets of distractor-initial trials.

	Looking RT	Looking proportion	Pointing accuracy
Looking RT	.	−.537	−.140
Looking proportion		.	.590
Pointing accuracy			.

values for the current dataset, limited to trials with valid RTs, is .27, 95% CI: [.20, .34] (see Supporting Information for more reliability data). Another interpretation is that greater response certainty when testing familiar words leads to different, more diagnostic looking RT patterns. Peekbank looking RT data from older children (see Discussion) showed especially low reliability for novel-word trials. A quick check of looking RTs on the familiar-word trials examined here showed a numerically stronger relationship with pointing ($r=-.345$, 95% CI: [−.518, −.144]), compared to novel words here ($r=-.199$, 95% CI: [−.284, −.110]). Though neither of these two observations is conclusive, both are consistent with looking RT responses being less robust for novel words. A final possibility is that something about the pointing task altered looking responses in some way that reduced the diagnosticity of looking RT. Future work might examine roles of age, word familiarity, and pointing task on looking RT measures.

Question 2: Do looks detect associative knowledge that pointing does not? Not really

One interpretation that one might derive from the developmental literature is that looking measures are more sensitive to partial or implicit knowledge than overt measures like pointing. If this is so, then looking measures should exceed chance even when pointing accuracy is at chance. One way to test this prediction is to examine the *intercept* of the earlier model predicting looks from pointing. Of course, this analysis presumes a linear relationship, which probably is not quite correct. Thus, a second test of whether looking reveals additional knowledge is to examine looking proportions for participants whose accuracy is at chance. I did both.

In a linear regression model using pointing accuracy to predict looks, when accuracy is at chance, do looks at the intercept exceed chance? By a trivially small amount

In the original model computed above, the intercept term exceeded chance ($B=.015$, $SE=.006$, $t=2.29$, $p=.02$), but only assuming the most generous alpha threshold (.05). Further, the beta of .015 is not large: it means looking at the target .431 and the other picture .416, or about 50.9% target looks. This calculation is based on the fact that on average, for usable trials, 84.8% (proportion = .848) of looks were to either target or competitor picture (other locations included the center of the screen, off-screen looks, or blinks).

Since some findings (reviewed in Introduction) suggest that looking-pointing relationships might be weaker at earlier ages, I considered individual age groups. Three-year-olds ($B=.020$, $SE=.009$, $t=2.28$, $p=.02$) and 4-year-olds ($B=.022$, $SE=.008$, $t=2.60$, $p=.009$) showed (unimpressively) significant intercept terms, but 5-year-olds did not ($B=-.008$, $SE=.013$, $t=-0.628$, $p=.51$). Nonzero intercepts at ages 3–4 but not at 5 years might happen if looking and pointing become more tightly coupled with age. Even this is an interpretive stretch, as the main effect of age did not reach significance, implying that intercept values at different ages were not significantly different.

Nonzero intercepts might also happen spuriously if the relationship between looks and points is not (quite) linear. Thus, the next analysis examines only the subset of participants whose pointing accuracy was numerically close to chance, eliminating the linear component.

Do participants around chance pointing accuracy show above-chance looking? No

I extracted looking data from cells with pointing accuracy between .45 and .55 (Figure 5), yielding 244 data points (229 participants; 15 participants contributed two

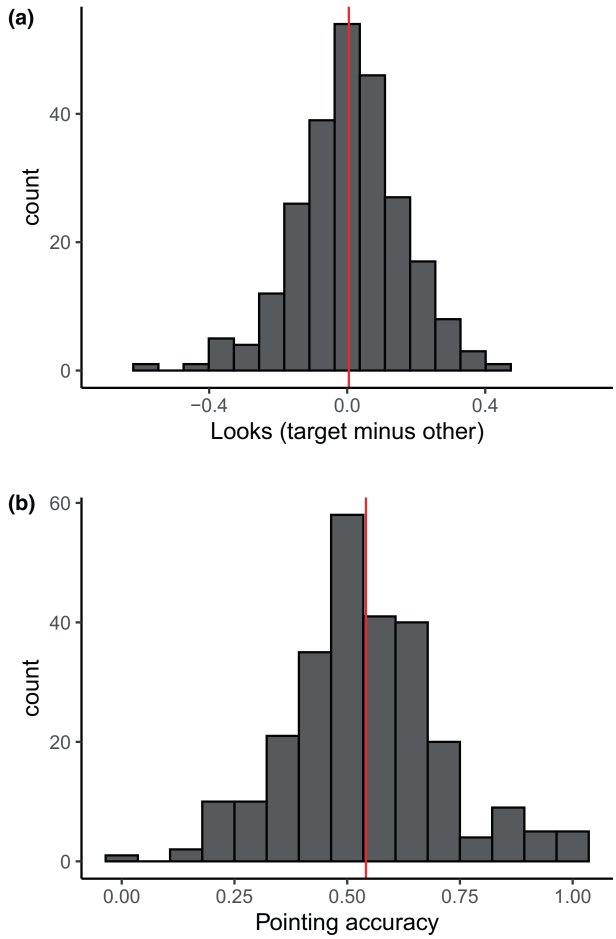

FIGURE 5 (a) Looking proportions when pointing is at chance (.45 ≤ pointing ≤ .55). (b) Pointing accuracy when looking proportion in 200–2000-ms window is at chance (−.05 ≤ looking ≤ .05). Vertical red lines indicate means.

data points). For comparisons to chance, I computed Bayesian t-tests in JASP version 0.17.2 (JASP Team, 2023), setting the alternative hypothesis to greater than chance. Conventionally, a Bayes factor (BF_{10}) over 3 is moderate evidence in favor of the alternative hypothesis, while a BF_{10} below 1/3 is moderate evidence in favor of the null hypothesis, that is, chance performance. Since data in this range might be unevenly distributed (e.g., if there were more data points above .50 than below .50), I tested the pointing accuracy score ($M = .4998 \pm .021$) against .50, and it was not significant ($t(243) = -0.18$, $p = .86$; nonparametric Wilcoxon signed rank test, $V = 1505$, $p = .99$; $BF_{10} = 0.063$), verifying that this subsample reflects a chance central tendency.

Next, to answer the critical question, whether there are above-chance looks for chance points, I tested the looking score ($M = .0047 \pm .151$) against 0; it too was not significant ($t(243) = 0.48$, $p = .63$; Wilcoxon test, $V = 15{,}810$, $p = .43$; $BF_{10} = .110$). Last, visual inspection (Figure S5) suggests that chance pointers do not show "good" looking patterns at some limited time region.

In short, chance pointers are also chance lookers

This suggests that, for newly learned associations, looks are not detecting knowledge that is absent from pointing accuracy, nor are they tapping a more implicit level of knowledge than overt pointing behavior.

Exploration: Does chance looking predict chance pointing? Not exactly

Intercept of linear model where looks predict points exceeds chance

One might expect that flipping the direction of prediction would generate an identical outcome. I created a linear regression model analogous to the model for Question 1, but where looks predict pointing (full results in Supporting Information). The intercept was significant ($B = .060$, $SE = .008$, $t = 7.97$, $p < .0001$) suggesting that chance lookers sometimes point above chance. The beta of 0.060 equates to pointing accuracy of 56% (6% above chance rate of 50%).

Pointing accuracy for at-chance lookers exceeds chance

I also examined pointing accuracy scores of the chance-looking participants (Figure 5, right). To obtain roughly the same number of participants as in the chance pointing analyses earlier, I selected looking scores spanning zero ($-.05 < \text{looks} < .05$), yielding 261 data points from 245 participants (16 participants contributed two values). To check for uneven distribution (e.g., if there were more data points above 0 than below 0), I verified that looks in this subsample did not exceed chance ($.001 \pm .028$; $t(260) = 0.64$, $p = .52$; Wilcoxon $V = 17{,}882$, $p = .5197$; $BF_{10} = 0.125$). However, supporting the intercept analyses, pointing accuracy in the same set of participants exceeds chance ($.542 \pm .170$; $t(260) = 3.94$, $p = .0001$; Wilcoxon $V = 13{,}992$, $p = .0006$).

How can chance lookers be accurate pointers? They seem to be late lookers rather than true chance lookers

Factors external to target selection (e.g., raw visual salience, familiarity, and preference) may cancel out or override looking patterns. An additional possibility is that the averaged looking measure here incompletely captures children's looking behavior, and they *are* looking above chance. This might happen if the high-pointing-accuracy "non-lookers" only look to targets after the end of the analysis window. To assess this possibility, I extracted "non-lookers" with high pointing accuracy (.75+; $n = 33$; 6.1%) and calculated their looking proportions over time to 3000 ms (past the end of the 200–2000 ms analysis window). Figure S6 suggests that their looks increase toward and after the end of the 200–2000 ms window. They are *late lookers* but not *non-lookers*. In any case, they represent only a small proportion (6.1%) of the 545 data cells with high pointing accuracy.

Exploration of too-early looks: Pre-signal gaze location relates to pointing response

One interesting feature of the looking-pointing correlations over time is that looks in the 200–300 ms time bin already correlate with pointing (see Figure 4 above). Further examination shows positive correlations even earlier than 200 ms (Figure S7), even though auditory event recognition is unlikely to affect looks before 200 (see Salverda et al., 2014, for a 200-ms lower limit on signal-driven looks in adults). That is, a child who eventually points to the target is more likely to have been looking at the target picture prior to auditory event-driven looks (dark blue lines), and the reverse for a child who points to the distractor (green lines). Fifty-three out of 60 conditions (binomial probability $p < .0001$), show this effect numerically (Figure S8).

One plausible reason for this non-zero relationship of early looks with eventual choice is that children's early looks reflect interest (image features, idiosyncratic preferences, and even moment-to-moment variations in attention), and interest biases the child's choice. Another possibility is that *happening to be looking at* something early on captures attention, and heightened attention itself contributes to pointing decisions. This effect appears to diminish among more accurate pointers (Figure S9), suggesting that higher certainty may decrease the likelihood of early gaze location affecting final pointing choices. Briefly, this exploratory analysis suggests that eliminating error trials from gaze data (on analogy with RT analyses) may give the false impression that participants look at the correct response before they have enough information to do so. I return to this point in the Discussion.

DISCUSSION

I examined the relationship between two dependent measures commonly used in developmental research: looking time (infants) and pointing (children) in data from over 900 2.5- to 6.5-year-olds who all provided both measures. I found strong associations between looking proportions and pointing accuracy. Associations were stronger with later and longer looking time windows, with pointing accuracy explaining up to 50% of the variance in looking proportions across the whole dataset. I also found that, in this age group, when learning new associations, chance pointers are also chance lookers. That is, counter to a "looks-first" hypothesis, looks do not appear to detect additional knowledge that pointing misses or that children cannot yet express via pointing.

Implications for development

Looks and points line up qualitatively

Overall, these findings verify strong relationships between looking and pointing measures in early childhood, at least by age 3 years. This implies potential developmental continuity between infant-looking measures and child-looking measures, and describes a potential path forward for assessing continuous developmental change. The best approach to examine continuous developmental change would be to use the same measure across age, such as looking proportions, which (unlike pointing) can be measured from a very early age. However, before doing so, it is crucial to establish the *validity* of looking proportions as a measure of recognition. The current data suggest that looking proportions are valid in that they are closely related to pointing responses. Still, further research should assess looking-pointing relationships in younger children, ideally down to an age where pointing becomes unreliable, to better bridge the empirical gap.

The current data provide no support for "looks first"

At least by age 3 years, looking and pointing may not reflect different levels of processing. That is, looking time measures in this age group do not appear to reflect a different "level" of processing than pointing that is sensitive to knowledge that pointing misses. If children point at chance, looks are correspondingly at chance. Finding this pattern for newly learned word-label pairings is especially interesting in that new associations might be construed as fragile knowledge compared to knowledge of familiar word-label pairs, and fragile knowledge is a case where looking measures sometimes outperform pointing measures (see review in Introduction). What is less clear is how far back in development this relationship holds, as discussed below.

It is noteworthy that looking patterns are more strongly related to pointing accuracy than to chronological age. This is consistent with a large common factor underlying both behaviors, which I would argue to be ease of recognition. The strong looking-pointing relationship suggests a new interpretation to existing findings of faster looking RTs and larger gaze proportions to named pictures with age (e.g., Fernald et al., 1998, 2006; Fernald & Marchman, 2012): Perhaps these measures reflect increasing knowledge *of particular words*, not just general verbal ability.

One question raised earlier was what it means when very young infants look a small amount at named targets. With the above discussion in mind, the current set of results simply implies that looking a little is rarely associated with reliable pointing responses. Cautiously, we might infer that looking a little means knowing a little. Thus, increases in target looking proportions with age or with greater word familiarity indicate stronger knowledge.

Further considerations and limitations

There are a number of unanswered questions here, centering on developmental change. The primary limitation is the absence of data at younger age ranges where children can still point. I am confident based on current data that 2.5-year-olds would show a strong looking-pointing relationship. I am less confident for younger ages. Hayne et al. (2016) found that children aged 1 years; 0 only pointed to named familiar pictures when prompted ("Where's Dad?") about 30% of the time. I am also less confident about children with variable language skills, such as those characterized as having developmental language disorder.

The strong looking-pointing relationship may only materialize at 3 years or so, with an advantage for looking over manual responses earlier in development and possibly also in atypical populations. Some word recognition evidence is consistent with an advantage for looking (vs. touching/pointing) in children under 2 years (Gurteen et al., 2011; Hendrickson et al., 2015, 2017). Still, Bakopoulou et al. (2023) found that in children around age 2 years, looking and object selection were strongly related, Hendrickson et al. (2015) found that 16-month-olds who touched more accurately for parent-reported "known" words also looked more quickly, and Gurteen et al. (2011) found that for familiar control words, looking and pointing were similar. Those researchers used different dependent measures than I did, precluding direct comparison. Nonetheless, those studies suggest exciting potential for examining looking and pointing (or other manual measures) jointly in younger children.

A relevant question is, if looking and pointing are less tightly coupled earlier in development, why so? One possibility already considered is that both looking and pointing access the same word and auditory knowledge but eye movements are less motorically costly, and thus appear earlier and more reliably during development. A second possibility is that the ability to control eye movements develops earlier than larger motor skills, and contributes to the development of manual movements by providing a supervisory signal. A final possibility is that looking and pointing may be nonidentical because looks are affected by partly nonoverlapping factors (e.g., visual salience, visual appeal, or novelty) whose effects may be strongest at the youngest ages (see, e.g., Newman et al., 2001).

Another consideration is that time course measures may be informative. Time course measures achieved by curve-fitting (slopes, trajectories, and polynomial fits) have not been considered here. This matches many infant studies of word recognition that use averaged looking time. Still, some characterization of time course may turn out to be more sensitive or more informative than the "area under the curve" approach used here (though see McMurray, 2023, for interesting simulations and a note of caution with regard to curve-fitting). It is worth noting that time course is related to total looking time: if a child fixates the correct picture sooner, they will have a sharper looking slope as well as a larger looking proportion. Still, for highly familiar words, where young children or even infants might be expected to have ceiling-level knowledge, rapidity of fixation or curve fits might be more informative correlates of ease of recognition, with the caution that curve-fitting measures tend to require a large number of trials to obtain stable estimates (see McMurray, 2023). Future work should assess the tradeoffs of curve-fitting versus area-under-the-curve measures across different situations (age of participants, level of manual response accuracy, and number of pictured alternatives). Developmentally, though, recognition of "familiar" words may *not* be at ceiling at very young ages, in which case, the current dataset of responses to newly learned words may be a reasonable analogy.

To put the current results in context, it is important to remember that looking patterns contain more information than simple recognition, including differences in temporal availability of information. While as much as 50% of variability may be explained by final pointing response choices, this leaves a lot of unexplained variability, particularly early in the trial. Other influences might include: idiosyncratic, fluctuating levels of interest in pictures; visual features like perceived size or contrast; spatial location within a display; general picture interestingness or semantic interestingness (e.g., delicious things, animate things, and people). Finally, differences in the temporal availability of information likely exert influences, yielding earlier looks with earlier information. This includes my own research in some cases (Creel & Tumlin, 2011, 2012), where voice characteristics versus speech sounds, or absolute pitch versus relative pitch, become available on different time scales.

A final factor that may influence looking patterns is the instruction to generate an action (pointing and reaching) in the first place. For example, when adults are simply asked to look at the display while hearing spoken material, rather than to select elements of the display, they show slightly less deterministic fixation proportions (e.g., Altmann & Kamide, 1999, Experiment 2) than when asked to evaluate the relevance of the spoken sentence to the scene (e.g., Altmann & Kamide, 1999, Experiment 1; see also Allopenna et al., 1998).

One might reasonably ask whether the looking-pointing correlation here is artificially inflated by the task itself: Perhaps children show less sustained duration of named-picture looks if they do not have to plan and execute a pointing response. In particular, pointing might cause children to sustain looks to the target later in the trial only in order to guide their pointing responses, rather than sustained target looks being a natural consequence of comprehension. If true, this would predict two different patterns of looking over time for looks-only versus looking-and-pointing studies (Figure 6, top). Looking-only studies should show a rise in looks to the named object, and then a fall back to baseline. Looking-and-pointing tasks, by contrast, should show a rise in looks to the named object and that rise should be sustained through the end of the trial.

To assess this possibility, I gathered empirical evidence of looking patterns from looking-only studies with children in Peekbank (Zettersten et al., 2023), an online repository of child eye-tracked language studies. I used peekbankr (Braginsky et al., 2024) to extract data. Six studies included 20 or more children in the same age range as this work (Frank et al., 2016; Pomper & Saffran, 2016, 2019; Yurovsky & Frank, 2015; Yurovsky et al., 2013, 2017). I verified with the authors of those studies that children were not instructed to point (M. Frank, R. Pomper, J. Saffran, D. Yurovsky, personal communications, April 18, 2024).

In brief, looking patterns for words in the current studies qualitatively resembled looking patterns from looks-only studies in Peekbank. In Figure 6 (lower), none of the Peekbank looking-only data (colors, thinner lines) exhibit the pattern in Figure 6 (upper), instead resembling the current data (black, thick lines). Analyses are detailed further in the Supporting Information. This

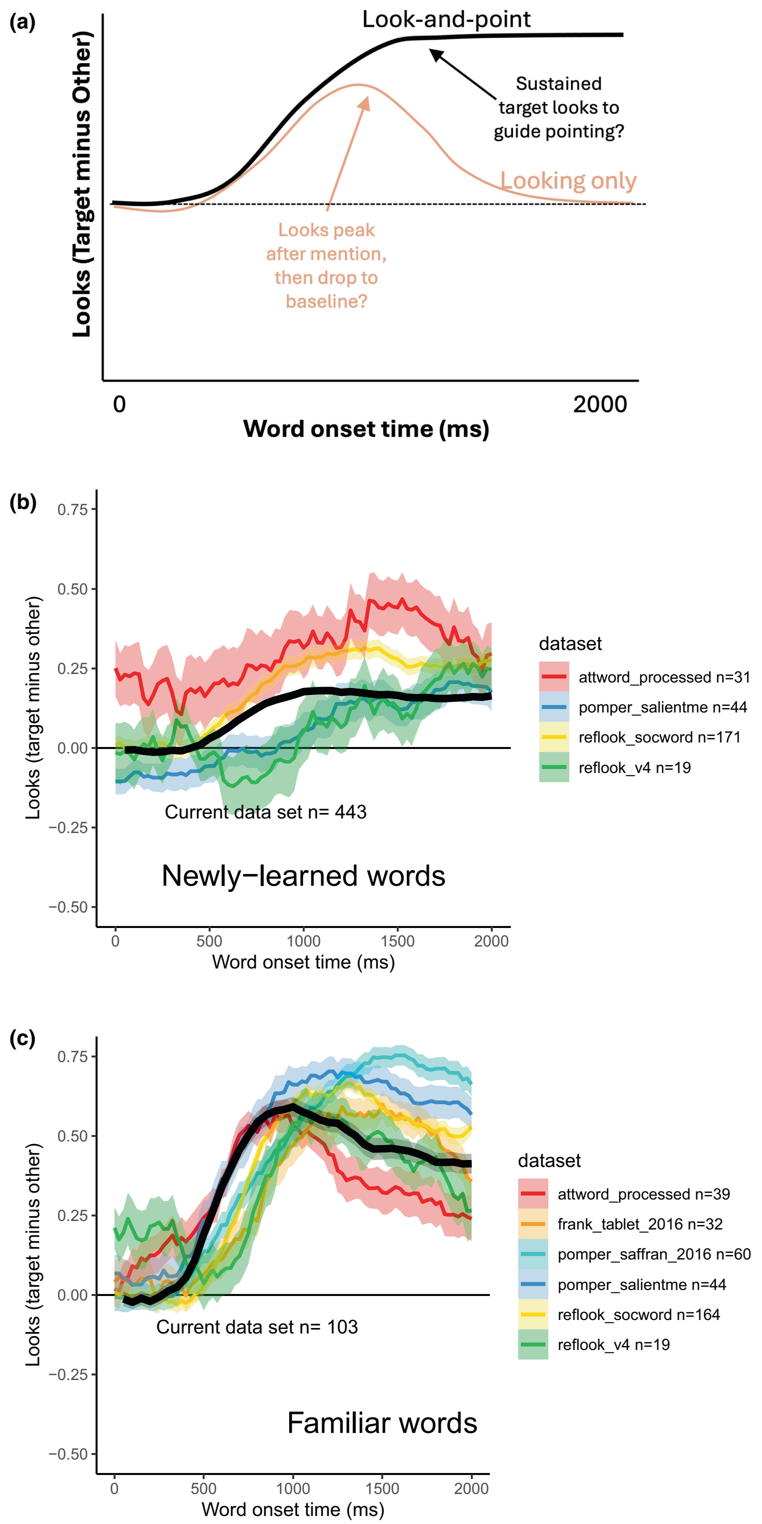

FIGURE 6 (a) Possible differences in looking pattern depending on whether or not children are asked to point. Looking patterns over time for word learning (b) and familiar-word (c) studies. Current data: Black, thick lines; Peekbank datasets: Thinner, colored lines.

does not support the hypothesis that the correlation between looking and pointing is an artifact of the pointing task. Future work should refine this observation by examining whether there are subtler effects of action planning and execution on children's looking patterns during word recognition.

Implications for interpretation and methodology

There are several clear implications for conducting visual-world eye tracking studies of learning in young children and, by extension, infants. First, looking proportion measures appear more closely related to pointing accuracy than looking RTs. While there are multiple possible reasons for this, the finding may recommend proportion measures over looking RT measures, at least at small trial numbers and when testing newly learned materials. An alternative approach might be to optimize for total number of usable looking RTs (e.g., see Adams et al., 2018, for high RT data output; Egger et al., 2020, for a gaze-contingent paradigm).

Second, considering looking proportion measures, researchers should leave in trials where an overt response is incorrect, for two reasons. One, the researcher risks getting spurious "precognitive" looks where children appear to look at the correct response prior to hearing the stimulus. Two, it makes results less comparable to studies from infants or toddlers who might make "errors" covertly, but whose errors cannot be omitted due to absence of overt responses.

Third, some time windows are better than others. A long time window is good for assessing recognition accuracy, assuming that is what the researcher wants to measure. There is no clear "cutoff" window that stands out as best, but very early looks are not strongly related to overt responses. The 200–2000-ms window that many researchers (including me) have used appears to correspond well to pointing responses. The 200- to 3000-ms window does also, and some infant researchers have utilized this or longer windows. However, extending the window to 3000 ms, at least in this age group, is nonideal in that between 2000 and 3000 ms, many trials are ending (nearly 40% have ended by 3000 ms), meaning children have long since pointed and their response has been entered by the experimenter. Further, the additional 1000 ms do not appear to result in substantial gains in the looking-pointing relationship.

CONCLUSION

I examined a large dataset of 2.5- to 6.5-year-old children learning words and other audiovisual associations and then both looking and pointing. Looking proportion measures and pointing responses are strongly coupled, suggesting potential developmental continuity between these two measures. Looking RTs in this data were less strongly related to pointing. I found almost no evidence that chance pointers betray additional knowledge in their looking behavior. This work represents a first step in characterizing continuity in the development of word recognition, with implications for examining developmental continuity in other domains.

ACKNOWLEDGMENTS

I thank Richard Aslin, Arielle Borovsky, Gedeon Deák, Shannon Ellis, and Martin Zettersten for comments on various versions of this manuscript. We also thank laboratory assistants over the years for supervising collection of the data included here: Alyssa Foote, Heather Pelton, Adrienne Moore, Dolly Rojo, Emilie Seubert, Nicolle Paullada, Liz Groves, and Alicia Escobedo, as well as Creel lab graduate students, particularly Reina Mizrahi and Conor Frye, and former honors student Sofia Jiménez. Finally, many thanks to the schools, teachers, parents, and children who made this work possible.

FUNDING INFORMATION

Data were collected under NSF grants BCS-1057080 and BCS-1230003 to SCC.

DATA AVAILABILITY STATEMENT

This study was not preregistered. Data and analysis scripts necessary to reproduce the analyses are publicly available at https://osf.io/wszd5/?view_only=59c42591acae41c3b890e66e4b8e98c0. Materials necessary to attempt to replicate the findings are available from the author.

ORCID

Sarah C. Creel https://orcid.org/0000-0002-1891-4228

REFERENCES

Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., & Feldman, H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes in full term and preterm toddlers. *Child Development*, *89*(5), 1674–1690. https://doi.org/10.1111/cdev.12818

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Bakopoulou, M., Lorenz, M. G., Forbes, S. H., Tremlin, R., Bates, J., & Samuelson, L. K. (2023). Vocabulary and automatic attention: The relation between novel words and gaze dynamics in noun generalization. *Developmental Science*, *26*(6), 1–9. https://doi.org/10.1111/desc.13399

Bergelson, E., & Aslin, R. (2017). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, *13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

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. (2014). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*, 369–380. https://doi.org/10.1080/15475441.2014.979387

Berthier, N. E., Sylvia, M. R., Johnson, R. L., & Clifton, R. K. (2001). Using object knowledge in visual tracking and reaching. *Infancy*, *2*(2), 257–284.

Braginsky, M., MacDonald, K., & Frank, M. (2024). *peekbankr: Accessing the Peekbank database*. R package version 0.1.1.9002. http://github.com/langcog/peekbankr

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2022). Six solutions for more reliable infant research. *Infant and Child Development*, *31*(5), 1–19. https://doi.org/10.1002/icd.2296

Cooper, R. M. (1974). The control of eye fixation by the meaning spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.
Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*(2), 239–264. https://doi.org/10.1177/1745691615621279
Creel, S. C. (2014a). Impossible to _gnore: Word-form inconsistency slows preschool children's word-learning. *Language Learning and Development*, *10*(1), 68–95. https://doi.org/10.1080/15475441.2013.803871
Creel, S. C. (2014b). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*, *73*, 81–98. https://doi.org/10.1016/j.jml.2014.03.001
Creel, S. C. (2014c). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1146–1160. https://doi.org/10.1037/a0036057
Creel, S. C. (2016). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science*, *40*, 373–403. https://doi.org/10.1111/cogs.12237
Creel, S. C., & Frye, C. I. (2024). Minimal gains for minimal pairs: Difficulty in learning similar-sounding words continues into preschool. *Journal of Experimental Child Psychology*, *240*, 105831. https://doi.org/10.1016/j.jecp.2023.105831
Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509.
Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in processing? *Trends in Cognitive Sciences*, *19*(12), 713–716. https://doi.org/10.1016/j.tics.2015.09.006
Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285. https://doi.org/10.1016/j.jml.2011.06.005
Creel, S. C., & Tumlin, M. A. (2012). Online recognition of music is influenced by relative and absolute pitch information. *Cognitive Science*, *36*(2), 224–260. https://doi.org/10.1111/j.1551-6709.2011.01206.x
Cuevas, K., & Bell, M. A. (2010). Developmental progression of looking and reaching performance on the a-not-b task. *Developmental Psychology*, *46*(5), 1363–1371. https://doi.org/10.1037/a0020185
Diamond, A. (1995). Evidence of robust recognition memory early in life even when assessed by reaching behavior. *Journal of Experimental Child Psychology*, *59*, 419–456.
Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352
Egger, J., Rowland, C. F., & Bergmann, C. (2020). Improving the robustness of infant lexical processing speed measures. *Behavior Research Methods*, *52*(5), 2188–2201. https://doi.org/10.3758/s13428-020-01385-5
Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x
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., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*(4), 1003–1015. https://doi.org/10.1111/1467-8624.00331
Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17. https://doi.org/10.1080/15248372.2015.1061528
Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly*, *46*(2), 342–369.
Frye, C. I., & Creel, S. C. (2022). Perceptual flexibility in word learning: Preschoolers learn words with speech sound variability. *Brain and Language*, *226*(January), 105078. https://doi.org/10.1016/j.bandl.2022.105078
Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217. https://doi.org/10.1016/j.jecp.2010.12.001
Hayne, H., Jaeger, K., Sonne, T., & Gross, J. (2016). Visual attention to meaningful stimuli by 1- to 3-year olds: Implications for the measurement of memory. *Developmental Psychobiology*, *58*(7), 808–816. https://doi.org/10.1002/dev.21455
Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121
Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science*, *18*(5), 723–735. https://doi.org/10.1111/desc.12250
Hendrickson, K., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2017). Assessing a continuum of lexical–semantic knowledge in the second year of life: A multimodal approach. *Journal of Experimental Child Psychology*, *158*, 95–111. https://doi.org/10.1016/j.jecp.2017.01.003
JASP Team. (2023). *JASP (Version 0.17.2)* [Computer software].
Jiménez, S. R., & Creel, S. C. (2012). Factors affecting talker recognition in preschoolers and adults. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University conference on language development* (pp. 246–257). Cascadilla Press.
Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, *12*(3), 79–83. https://doi.org/10.1111/1467-8721.01234
Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(1992), 1311–1328. https://doi.org/10.1068/p2935
Lee, V., & Kuhlmeier, V. A. (2014). Young children show a dissociation in looking and pointing behavior in falling events. *Cognitive Development*, *28*(1), 21–30. https://doi.org/10.1016/j.cogdev.2012.06.001
Marchman, V. A., Ashland, M. D., Loi, E. C., Munevar, M., Shannon, K. A., Fernald, A., & Feldman, H. M. (2022). Associations between early efficiency in language processing and language and cognitive outcomes in children born full term and preterm: Similarities and differences. *Child Neuropsychology*, *29*(6), 886–905. https://doi.org/10.1080/09297049.2022.2138304
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), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x
McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*, 631.
McMurray, B. (2023). I'm not sure that curve means what you think it means: Toward a [more] realistic understanding of the role of eye-movement generation in the Visual World Paradigm. *Psychonomic Bulletin and Review*, *30*(1), 102–146. https://doi.org/10.3758/s13423-022-02143-8

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203–229. https://doi.org/10.1207/s15327078in0602_4

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52–61. https://doi.org/10.1016/j.jecp.2011.03.014

Newman, C., Atkinson, J., & Braddick, O. (2001). The development of reaching and looking preferences in infants to objects of different sizes. *Developmental Psychology*, *37*(4), 561–572. https://doi.org/10.1037//0012-1649.37.4.561

Parsons, S. (2021). splithalf: Robust estimates of split half reliability. *Journal of Open Source Software*, *6*(60), 3041. https://doi.org/10.21105/joss.03041

Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, *2*(4), 378–395. https://doi.org/10.1177/2515245919879695

Pomper, R., & Saffran, J. R. (2016). Roses are red, socks are blue: Switching dimensions disrupts young children's language comprehension. *PLoS One*, *11*(6), 1–19. https://doi.org/10.1371/journal.pone.0158459

Pomper, R., & Saffran, J. R. (2019). Familiar object salience affects novel word learning. *Child Development*, *90*(2), e246–e262. https://doi.org/10.1111/cdev.13053

Quam, C., & Swingley, D. (2010). Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, *62*(2), 135–150. https://doi.org/10.1016/j.jml.2009.09.003

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250. https://doi.org/10.1111/j.1467-8624.2011.01700.x

Quam, C., & Swingley, D. (2014). Processing of lexical stress cues by young children. *Journal of Experimental Child Psychology*, *123*, 73–89. https://doi.org/10.1016/j.jecp.2014.01.010

Quam, C., & Swingley, D. (2022). A protracted developmental trajectory for English-learning children's detection of consonant mispronunciations in newly learned words. *Language Acquisition*, *30*, 256–276. https://doi.org/10.1080/10489223.2022.2069026

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rohatgi, A. (2019). *WebPlotDigitizer version 4.2.0: A web based tool to extract numerical data from plot images* [computer program]. https://automeris.io/v4/

Salverda, A. P., Kleinschmidt, D. F., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Smolak, E., Hendrickson, K., Zesiger, P., Poulin-Dubois, D., & Friend, M. (2021). Visual and haptic responses as measures of word comprehension and speed of processing in toddlers: Relative predictive utility. *Journal of Experimental Child Psychology*, *203*, 105032. https://doi.org/10.1016/j.jecp.2020.105032

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1321–1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1312–1323. https://doi.org/10.1044/1092-4388(2003/102)

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

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

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

Vishton, P. M., Ware, E. A., & Badger, A. N. (2005). Different gestalt processing for different actions? Comparing object-directed reaching and looking time measures. *Journal of Experimental Child Psychology*, *90*, 89–113. https://doi.org/10.1016/j.jecp.2004.10.002

Von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Tracking and reaching for moving objects. *Cognition*, *67*, 255–285.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–418. https://doi.org/10.1111/j.1467-7687.2008.00789.x

Yurovsky, D., & Frank, M. C. (2015). Beyond naive cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*, e12349. https://doi.org/10.1111/desc.12349

Yurovsky, D., Wade, A., & Frank, M. C. (2013). Online processing of speech and social information in early word learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 1641–1646). Cognitive Science Society.

Yurovsky, D., Wade, A., Kraus, A. M., Gengoux, G. W., Antonio, Y., & Frank, M. C. (2017). *Developmental changes in the speed of social attention in early word learning* [unpublished manuscript].

Zettersten, M., Yurovsky, D., Xu, T. L., Uner, S., Tsui, A. S. M., Schneider, R. M., Saleh, A. N., Meylan, S. C., Marchman, V. A., Mankewitz, J., MacDonald, K., Long, B., Lewis, M., Kachergis, G., Handa, K., deMayo, B., Carstensen, A., Braginsky, M., Boyce, V., … Frank, M. C. (2023). Peekbank: An open, large-scale repository for developmental eye-tracking data of children's word recognition. *Behavior Research Methods*, *55*(5), 2485–2500. https://doi.org/10.3758/s13428-022-01906-4

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Creel, S. C. (2024). Connecting the tots: Strong looking-pointing correlations in preschoolers' word learning and implications for continuity in language development. *Child Development*, *00*, 1–17. https://doi.org/10.1111/cdev.14157