Selective Attention in Preschoolers Based on Speech Complexity and Learning Rate

Abstract

We introduce a novel method to test a classic idea in developmental science: that children's attention to a stimulus is driven by how much they can learn from it. Preschoolers (4–6 years, $M = 4.6$) watched a video where a distracting animation accompanied static, page-by-page illustrations of a storybook. The audio narration for each storybook page was looped such that children could listen to it up to 6 total times. However, the narration automatically ended if the child looked at the distractor for an extended period of time, indicating their loss of attention to the story, and triggering the next page. The complexity of the narration was manipulated between-subjects: the Simple narration largely contained words that should be familiar to preschoolers, while the Complex narration contained many rare, late-acquired words. Children's learning was measured via post-tests of their plot comprehension and knowledge of the rare words. Consistent with the hypothesis that children's attention was driven at least partly by their ability to learn from the speech, we observed a significant interaction between complexity level and age in predicting children's probability of continuing listening on each page. That is, while younger children were more likely to continue listening to the Simple speech, older children became increasingly likely to sustain attention to the Complex speech. Our results provide evidence that young children may actively direct their attention toward linguistic input that is most appropriate for their current level of cognitive and linguistic development, which may provide the best learning opportunities.

Keywords: selective attention, lexical development, self-directed learning, cognitive development, language processing, rational learning

Selective Attention in Preschoolers Based on Speech Complexity and Learning Rate

If you have ever read a young child a bedtime story, you have likely noticed how children will demand that you read some books over and over again, yet insist that you abandon others just as you begin. Moreover, the same—relatively simple—book that a toddler demands over and over may bore a preschooler, whose favorite—comparatively complex—book the toddler immediately rejects. This everyday example is suggestive of a general principle according to which children's attention is most readily sustained by information that they are best able to learn from: the toddler may have the sense that they are still learning from repeated narrations of the book that they favor, while the preschooler's favorite book is too far beyond the toddler's linguistic and world knowledge to readily support learning, leading to its immediate rejection. Notably, this hypothesis—that children's attention to an information source is driven by the degree to which it supports their learning—has its roots in foundational theory in developmental psychology (Bruner, 1961; e.g., Vygotsky, Cole, John-Steiner, Scribner, & Souberman, 1978), but has been difficult to obtain direct evidence for. Here, we employ a novel method inspired by the above scenario. Our study manipulates the complexity of a naturalistic speech stream and explores how children's attention to and learning from that speech shifts across a two-year age range, as children's linguistic competence and world knowledge grow. Support for this hypothesis would suggest a way in which children are *active learners*—and active *language* learners in particular—selectively attending to sources of linguistic information that they are best able to learn from.

Background

Previous work with infants provides evidence consistent with the idea that children's attention to a stimulus is driven by their sense of learning from it. In one body of work, for example, researchers independently define the complexity of different stimuli—irrespective of participants' knowledge or experience—and show that the duration of participants' attention systematically varies in response (Caron & Caron, 1969; Kidd, Piantadosi, & Aslin, 2012, 2014; Martin, 1975; Thomas, 1965). Many of these studies manipulate the predictability of highly simplified visual sequences, and use an ideal learner model to quantify the complexity of each event in the sequence via its surprisal. In an influential 2012 study, for example, Kidd and colleagues played simple sequences of visual events for 8-month-old infants, and measured infants' duration of attention in response. The authors dubbed the pattern they observed the ''Goldilocks effect:'' infants' probability of terminating attention was lowest for events of *intermediate* (or "just right") complexity (see also Kidd et al., 2014). Attending to intermediate levels of complexity is consistent with attending on the basis of learning, because the space between highly familiar and unmanageably novel is where learning is likely to be the most efficient. Most important for our purposes is that this 'U-shaped' relation between stimulus complexity and infants' probability of looking away from the display was evident not just at the group level, but in the habituation times of individual infants, at different ranges along the complexity continuum. This is the pattern that we would expect if infants' attention were driven by their sense of learning, because different ranges will be appropriate for different infants.

However, while attending to intermediate complexity is understood as a domain-general learning mechanism, studies showing complexity-based attention preferences are typically not designed to directly demonstrate the learning outcomes of early selective attention, leaving open the possibility that infants' attention reflects something more like a heuristic ("attend to medium complexity") rather than a responsive monitoring process ("attend while learning").

Studies that take us a step closer toward linking selective attention and learning are those that show how individuals' attention shifts with experience. For example, Forest et al. show how the complexity range of sequential visual stimuli that most attracts adults' attention advances as adults gain more experience with the stimuli. Poli et al. (2020) designed their experimental method to be able to link infants' attention to their learning progress: 8-month-olds watched individually-cued target shapes reappear at different locations on the screen. Each shape had a most-likely target location, making it possible to define learning progress via the information gain offered an ideal learner toward being able to predict the most likely target location for each shape. Infants' gaze in this paradigm did show the established relationship between complexity and attention, but learning progress proved an even stronger predictor of their habituation times. Not only that, but infants' actual learning progress was evident in their gaze as well, in that infants became faster and faster at directing their gaze toward predictable targets, consistent with having developed an efficient model of their statistical environment. Together, these studies show how learners' attention is (1) informed by relative complexity, and (2) how relative complexity is a moving target, informed by what learners have already seen—and it is notably almost always "seen."

Compatible with their stripped-down visual-event stimuli, these studies also employ a very specific notion of complexity as objective predictability, and operationalize learning as prediction. We are interested here in investigating these processes for learning beyond sequential statistical dependencies, and in particular for the higher-order sense-making involved in language comprehension at older ages. Triangulating on this idea, Kidd et al. (2014) extend the Goldilocks effect to auditory attention in infants, and, separately, find evidence of the effect in the visual attention of children ages 3 to 6. To our knowledge, one study tests these ideas with linguistic stimuli, narrowing in on the hypothesis that infants attend more to information that they would be more likely to learn from: Gerken and colleagues (2011) exposed 17-month-old infants to artificial language stimuli that either reflected an unlearnable grammatical pattern, or a grammatical pattern that infants of the same age in a prior experiment had been able to learn and generalize. Interestingly, infants took longer to habituate to the stimuli containing the latter, *subjectively learnable* grammatical pattern, leading the authors to propose a causal relation between so-called 'learnability' and attention. By this account, infants may implicitly monitor their own rates of learning from a particular information source, and disattend when their learning rate is below some threshold of efficiency. Notably, while learning is implicated as the underlying motivation for children's attention, these and other studies have not directly tested children's learning from the same stimuli to which attention is measured. They have also been limited in their capacity to say anything about learning by only varying the complexity of the stimulus, but not the relative competence of the learner. Subjective learnability is the product of the interaction of stimulus complexity and the relative competence of the learner. Thus, any study that only varies stimulus complexity cannot be sure that attentional preferences exhibited by children at the same level of development are a result of *learnability*, rather than irrelevant dimensions of the stimuli. We address these gaps using a novel paradigm and natural language stimuli.

The Present Study

The current study addresses these gaps and tests the hypothesis that children's attention to a source of linguistic information is driven by the degree to which it supports their learning. In a departure from previous studies employing highly simplified visual or auditory stimuli Kidd et al. (2014), we use natural language stimuli, which both interests children and carries real information for learning. Children across a two-year age range (4–6 years) listened to one of two alternate tellings of the same story, narrated at distinct levels of complexity: while the Simple story mostly used words that children are likely to know, the Complex story contained many words that were likely to be unfamiliar. During the story narration, we measured children's attention to the speech, and after the narration, we measured children's learning from the speech.

In thinking about our predictions, it is useful to distinguish between a child's sense that they are or could be learning something, and the product(s) of a child's learning—their learning *outcomes*. We expect the amount a child learns from a stimulus to be related to the amount they attended to it; that is, children's learning outcomes in our study and their attention to the story should be correlated. The more nuanced hypothesis that our study allows us to test is that a child's attention allocation is itself determined by the interaction between the complexity of a stimulus and the child's own competence. Here, we use child age as a proxy for linguistic competence. We expect that there will be a larger gap between the Complex speech and the language that the younger children in our sample know, and we expect that this gap will be smaller for older children. Thus, when listening to the Complex narration, we predict greater attention from the older children than from the younger children. Conversely, when listening to the Simple narration, we predict greater attention from the younger children than from the older children.

To begin to test these predictions, we played children in our study either a Simple or Complex narration of a textless storybook (Mayer, 1969) while their visual attention to a display was captured via an eyetracker, and directly tested children's learning outcomes as a result of the narration. On each page of the story, a continuously animated distractor (three penguins double-dutching) was presented alongside the static story illustration, thus competing for children's visual attention (Figure 1). Given the presence of this dynamic distractor, we reasoned that visual attention to the comparatively dull illustration was likely to be a meaningful index of children's attention to the speech. That is, we expected that children would continue to look at the static Illustration only as long as they were actively processing the story narration (even, that it would be difficult for them *not* to, as when the secret location of a queried object is unintentionally revealed by a child's gaze; Salverda and Altmann (2011); Cooper (1974)). Indeed, during piloting, children looked almost exclusively at the Distractor animation, rather than the Illustration, when the display was presented without the narration. When children were no longer listening to the story, we expected that they then might be lured by the DISTRACTOR. The duration of each storybook page was contingent on children's allocation of visual attention. Children who were consistently drawn in by the DISTRACTOR moved through the story quickly and heard the narration for each page only once. In contrast, children who continually gazed at the ILLUSTRATION (suggesting that they were paying attention to the speech) could hear each page of the story repeated up to five times (inset Figure 1). This method ensured that children heard the entirety of the story content (i.e., at least one repetition of each page) and provided attentional data for every storybook page, similar to the trial structure of gaze-contingent paradigms used with infants. Specifically, we quantify children's attention to the story by analyzing (1) whether children continued listening to further, optional repetitions of the narration for each page, and (2) how much they looked to the ILLUSTRATION, which was only made salient by the narration, rather than to the otherwise-salient DISTRACTOR.

To probe links between individual children's attention to the speech and their learning, we measured two learning outcomes after the story: (1) children's listening comprehension (their recollection of the content of the story), and (2; only for the children hearing the COMPLEX narration) their partial word knowledge of the rare target words embedded in the speech. These variables armed us to answer the following specific research questions: (1) is preschool-aged children's attention responsive to the complexity of naturalistic speech? (2) does children's age—as a proxy for their level of language and cognitive development—interact with experimentally manipulated complexity in predicting children's attention, consistent with children's attention being sensitive to the speech's support for their learning? and (3) within each condition, are children's attention to the speech and their learning outcomes correlated?

Materials and Methods

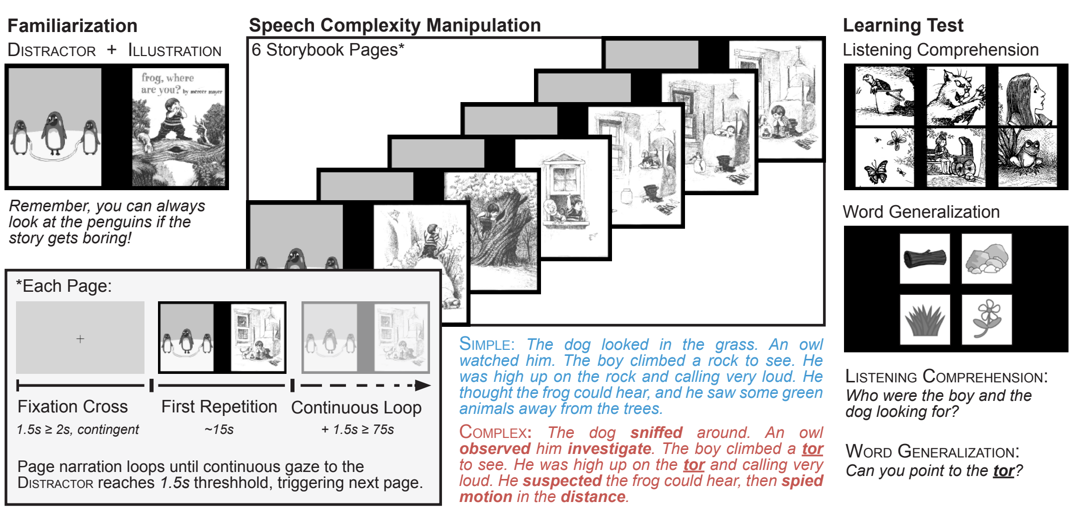


Figure 1. Schematic of Experimental Eyetracking Procedure Manipulating the Speech Complexity of a Narrated Storybook and Measuring Child Attention and Learning.

Full documentation of our procedures, including study scripts and stimuli, videos of experimental sessions, data analysis files, and supplementary information, is at [https://osf.io/zsjfb/?view_only=024c8e83e56a4fff95e5d5ae840035c2].

Participants

Our participants were forty-six children (4.0–6.0 years; $M = 4.61[0.13, 0.14]$, $SD = 0.47$) whose parents reported English as their primary language. Children came from a range of socioeconomic positions (17% with reported annual household incomes below 25K, 25% above 200K), with a skew toward higher-income households (50% of children came from households reporting 100K or more in annual income). Caregivers were overwhelmingly educated, with 75% of caregivers holding a graduate degree (only 17% of caregivers had completed fewer than 4 years of college). Children were generally identified by caregivers as Asian or Pacific Islander (42%) or White (42%), with 9% of children identified as Black, and 17% of children identified as belonging to multiple racial categories. Children were recruited from local preschools or from a database of interested families maintained by [removed to preserve anonymity], and tested in a quiet area of their school or in lab. Children received a sticker and/or certificate and small toy for their participation.

The COVID-19 pandemic forced us to halt data collection before reaching our planned sample of 64 children. However, a sensitivity analysis suggests that this nevertheless leaves us with over 80% power to detect a small crossover interaction between condition and age.

Prior to their study session, children were randomly assigned to the Simple ($n = 24$, $M = 4.61[0.20, 0.24]$, $SD = 0.54$) or Complex ($n = 22$, $M = 4.62[0.16, 0.18]$, $SD = 0.41$) condition. There was no significant difference between the ages of the children in the Simple and Complex conditions ($t(42.63) = -0.07, p = 0.946$). Two additional children were excluded after another child (1) or teacher (1) intervened on their study session.

Vocabulary Survey. To validate our assumptions about the words likely to be familiar versus unfamiliar to the children in our sample, we asked caregivers to fill out a vocabulary questionnaire. For every content word used in either condition of the study, caregivers indicated whether or not their child would "understand the word if you said it out loud." Caregivers typically filled out this measure, along with a demographic survey and language environment questionnaire, while the child was participating in the study. Caregivers of children tested in preschool were sent the link to the questionnaires over e-mail. This survey confirmed that the rare target words embedded in the complex condition were indeed novel to chidren: 0% of caregivers reported that they were familiar to their children (and many verbally reported having learned at least some of the words from the study themselves).

Procedure

Familiarization. Children sat before a laptop connected to an SMI RED-n eyetracker, wearing child-sized over-ear headphones. After a brief four-point calibration of the eyetracker ("Can you follow the little fairy on the screen?"), the familiarization began. The first screen displayed a black-and-white animation of three penguins jumping rope (the Distractor) on the left side of the screen (Figure 1). This screen lasted for 10s, during which a female voice drew the child's attention to the ongoing animation, and encouraged them to look there "if the story gets boring." Next, the cover of the book ("Frog, Where are You?"; Mayer (1969)) appeared alongside the Distractor. Both images were displayed for 15s, during which the voiceover reiterated that the child was going to hear a story, and again directed the child's attention to the Distractor ("Where are you going to look if the story gets boring?"). The familiarization phase ended with a looming fixation cross on a grey background, used to center children's gaze before the onset of the narration—and critical data collection—phase.

Storybook Narration. The same female voice narrated a boy and a dog's search for their escaped pet frog across six pages of a textless picture book. On each page, the Illustration for the story appeared on the right side of the screen, while the Distractor played continuously on the left. To ensure high-quality eyetracking data, a gaze-contingent fixation cross appeared between each page.

Speech Complexity Manipulation. Depending on the condition to which they were assigned, children heard the story narrated at either the Simple or Complex level (Figure 1). The Simple and Complex narrations were matched on multiple linguistic dimensions, but differed in the estimated age of acquisition (AoA) of the words they used (Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012).[1] The Simple narration exclusively used words from the MacArthur-Bates Communicative Development Inventory (Fenson et al., 2007), which is normed for children between 16 and 30 months. In contrast, each page of the Complex narration included five words with AoAs estimated between 7 and 13 years (bolded in the sample page narration in Figure 1), as well as a single rare and unfamiliar word with an estimated AoA of over 13 years, which was presented twice (bolded and underlined in Figure 1). The rare words were *ogled*, *absconded*, *flummoxed*, *hyaline*, *aperture*, and *tor* (two verbs, two adjectives, and two nouns). Children's learning of these rare words was assessed in the test phase.

Child-Controlled Listening. Children obligatorily heard the narration for each page at least once ($\sim 15s$), after which the same audio continued to loop for up to five additional repetitions ($\sim 75s$), separated by a $500 - 750ms$ pause. Children could advance to the next page early by looking at the Distractor: a fixation of 1.5s (1500ms) to the Distractor automatically triggered the next page. The child-controlled portion of the experiment lasted between 2s and 7s ($M = 3[3, 3]$).

Happy Ending. Regardless of condition, all children experienced the same (brief: 7s) end of the story: instead of the Distractor-Illustration split-screen, the display

[1] Across pages, narrations were matched for syllable count ($range : 50 - 61, M = 54.92[53.33, 56.59]$; $t(5.00) = 0.15, p = 0.889$ paired by page), speech rate ($range : 3.42 - 3.99, M = 3.67[3.58, 3.76]$; $t(5.00) = -0.16, p = 0.877$ paired by page), number of sentences (5/page) and number of questions vs. declarative sentences on each page. Sentences 1, 2, and 5 on each page—where the Complex narration embedded five later-acquired content words—were additionally matched on type-token ratio ($range : 0.81 - 1, M = 0.9[0.87, 0.94]$; $t(5.00) = -1.49, p = 0.197$ paired by page). Sentences 3 and 4 in the Complex condition used the rare target word for that page one time each.

showed facing storybook pages. The pages turned as the narrator described the boy and the dog's rediscovery of the frog (on a log surrounded by 'his whole family!').

Learning Tests. After the story, we measured children's learning outcomes via two blocks — Listening Comprehension and Unfamiliar Word Generalization — of six test trials each (one for each content page of the storybook). Three initial trials familiarized children with the format of the test questions, by asking them to point to the "dog," "boy," and "frog" in successive arrays. All children got these questions right. The subsequent test questions were always presented in the same order across children, within each block mirroring the pages on which the relevant information was introduced.

Listening Comprehension. In the first test block, *Listening Comprehension* trials tested children's knowledge of story events or characters. On each trial, the same narrator's voice asked a question (e.g., "Who were the boy and the dog looking for?") over a grey screen with a central fixation cross. When the child fixated on the cross, the screen switched to a 2x3 grid of black-and-white images (illustrations by the author-illustrator of "Frog, Where Are You?"; see the rightmost column of Figure 1). Children responded by pointing to one of the images.

Unfamiliar Word Generalization. *Unfamiliar Word Generalization* trials asked children in the Complex condition to generalize the unfamiliar target words that they had heard in the Complex narration to novel stimuli (e.g., from the boy 'ogling' the frog to a person peering through a magnifying glass, or from the frog in the story 'absconding' from the jar to a stylized graphic of a person running away).[2] As in the previous block of trials, children heard each test question (e.g., "Can you point to the person who is *absconding*?") over a grey screen with a central fixation cross. When children's fixation on the cross triggered the next screen, they responded by pointing to one of four candidate black-and-white illustrations, arranged in a 2x2 grid. Competitor images were selected to be compatible with the syntax of the test question (e.g., depicting other actions with thematic patients for 'ogling'), and the correct response for all questions was normed via a sample of undergraduates exposed to the same story narration ($N = 19$).

Variable Coding and Predictions

Child Attention Metrics. We captured variability in children's attention to the speech via measurements of: (1) children's probability of continuing listening beyond the first obligatory narration of each page, and (2) the duration and (3) the distribution of children's visual attention to our predefined Areas of Interest (AOIs).

Continued Listening. On each page, we coded whether the child moved on to the next page as soon as they could (that is, as soon as the obligatory first repetition of the narration for that page was over, plus the 1500ms threshhold for the trigger AOI: *continued listening = 0*), or continued listening for any amount of time past that (*continued listening = 1*). Coding children's listening time data in this way enabled us to

[2] Children in both conditions responded to these trials; however we only analyze data from children in the Complex condition, who actually heard the words in the story.

meaningfully analyze children's voluntary exposure to the speech, in spite of the challenges presented by children's raw listening durations after the first repetition of each page (namely, zero-inflation—many children moved on to the next page shortly after the first repetition—and a long tail; see *Supplementary Information*). Children moved on immediately on about a third of trials (29), and listened to all five additional repetitions on just five trials (1.81%).

While at the trial level, continued listening is a binary variable, at the subject level, we analyze it in terms of the proportion of pages on which a child continued listening (their *continued listening proportion*; $range: 0-1 M = 0.71[0.64, 0.78]$ across children). We take this measure to reflect a child's ongoing attention to the speech, or their sustained appetite for listening to more of it. Conversely, we can think of preschoolers' probability of 'moving on' from a storybook page in the present study paradigm as analogous to infants' probability of looking away in previous research.

Gaze to the Illustration vs. Distractor. For a more granular view of children's attention while listening to the story, we analyze continuous measures of their gaze to the two equal-sized AOIs we defined on the eyetracking display: the ILLUSTRATION and the DISTRACTOR.

Net Gaze Duration. A child's *net gaze duration* to a given AOI reflects the total time (in *ms*)[3] during which a child's gaze was both detectable by the eyetracker and fixated on that AOI. Thus, this measure combines information about the distribution of a child's attention during the story (i.e., between AOIs) *and* the overall length of their exposure to the story. At the trial level, each child contributes twelve net gaze duration data points: one duration value for each of the two AOIs, on each of the six storybook pages (ILLUSTRATION: $range: 0-62.78s, M = 15.42s[14.38, 16.44]$; DISTRACTOR: $range: 0-24.51s, M = 6.59s[6.12, 7.04]$). When analyzing net gaze duration at the subject level, we sum net gaze durations to the ILLUSTRATION across pages, and take a child's *total ILLUSTRATION gaze duration* as a global index of their attention to the speech ($range: 27.38-148.47s, M = 90.09s[82.3, 97.92]$; *total DISTRACTOR gaze duration*: $range: 14.11-73.11s, M = 39.54s[35.79, 43.47]$).

Percent Gaze Duration. A child's *percent gaze duration* for a given AOI represents their gaze to that AOI as a percentage of their gaze across the entire display.[4]

This measure narrows in on the relative *share* of children's visual attention devoted to each AOI (ILLUSTRATION: $range: 0-86s, M = 47s[45, 50]$; DISTRACTOR: $range: 0-94\%, M = 35\%[33, 38]$), irrespective of overall duration. As a trial-level index of attention to the story, we analyze children's percent gaze durations to the ILLUSTRATION, which we average across pages for a subject-level metric (*mean ILLUSTRATION percent gaze duration*: $range: 18-71\%, M = 47\%[43, 51]$; *mean DISTRACTOR percent gaze duration*: $range: 7-77\%, M = 35\%[31, 40]$).

[3] While we report descriptive statistics for gaze durations in *seconds* for readability (Table **??**), we use log-transformed *ms* values in our statistical models.

[4] The majority of children's gaze to the display—$M = 83\%[81, 85]$ was typically captured by one of our two AOIs.

If children's degree of attention to the speech is related to how appropriate it is for their current level of cognitive-linguistic competence, we should see an interaction between speech complexity and age in predicting children's attention. To illustrate with our 'continued listening' variable: in the SIMPLE condition, we might expect older children to typically move on from each page after hearing it once and likely extracting its information. On the other hand, we might expect younger children—who might still be learning from each SIMPLE page narration by the end of its first repetition—to be more likely to continue listening. In the COMPLEX condition, by contrast, we might expect children in this younger age group to have already disattended by the end of the first page repetition (because the complexity of the speech makes it difficult for them to learn from), and had their attention captured by the DISTRACTOR, causing the story to quickly advance to the next page. At the same time, we might expect *older* children — who have more hope of 'getting something' out of the more complex speech — to be more likely to *continue* listening past the first repetition of the page.

Learning Outcome Variables. We consider two measures of how well children were able to learn from the speech, one (*listening comprehension*) analyzable across all children, and the other (*unfamiliar word generalization*) applicable only to children in the COMPLEX condition.

Listening Comprehension. Children's responses on the six trials testing their knowledge of the story content were coded as correct (1) or incorrect (0). Children typically answered at least half of the questions correctly $(range : 0 - 100\%, M = 68\%[61, 75])$.

Unfamilar Word Generalization. Children's responses on the six trials testing their ability to generalize the rare target words embedded in the COMPLEX story narration were likewise coded as correct (1) or incorrect (0). Children showed highly variable performance on this measure $(range : 0 - 83\% accuracybychildinthe\textsc{Complex}condition, M = 39\%[31, 48])$. We analyze unfamiliar word generalization accuracy only in children in the COMPLEX condition, where children actually had the opportunity to learn something about the words during the experiment.

If children's attention is at least partly sustained by their sense that they are or could be learning something, we expect these measures of individual children's learning outcomes — evidence that they indeed *did* learn from the speech — to correlate with the above measures of their attention to the speech.

Analysis

In addition to reporting descriptive statistics regarding our variables of interest, we conduct two primary varieties of analyses: analyses testing the link between speech complexity and child attention, and analyses testing the link between child attention and learning outcomes.

In the first set of analyses, we use the lme4 package in R Douglas Bates, Mächler, Bolker, & Walker (2015) to fit mixed effects models (D. Bates, Mächler, B., & Walker, 2015) to the trial-by-trial data for each positive attention metric, separately (*continued* *listening*, ILLUSTRATION *net gaze duration*, and ILLUSTRATION *percent gaze duration*). We include condition (SIMPLE/COMPLEX), child age (mean-centered, in years), and the interaction of condition and child age as fixed effects, and random intercepts for child and page. In the case where a model of this structure fails to converge, we refit the model after dropping the random effect with the lowest variance (Barr, Levy, Scheepers, & Tily, 2013).

In the second set of analyses, we use mixed effects logit models to predict children's trial-by-trial test question accuracy, separately for each positive subject-level attention metric (*continued listening proportion*, *total* ILLUSTRATION *net gaze duration*, and *mean* ILLUSTRATION *percent gaze duration*), controlling for condition (listening comprehension only) and age (listening comprehension and unfamiliar word generalization). We standardize these variables to enable us to compare effect sizes across predictors.

We rely on model coefficients and odds ratios (ORs) to interpret the impact of different predictors on the dependent variables in each analysis, and assess the significance of individual predictors by comparing nested models with and without the relevant predictor (using the anova function in R; *A Language and Environment for Statistical Computing* (2020)).

Results

Is Children's Attention Responsive to Spoken Language Complexity?

Do children differentially attend to the Simple vs. Complex speech?

Children continued listening on an average of $M = 4.62[4.00, 5.21]$ $(range : 1.00 - 6.00)$ pages in the SIMPLE condition, and $M = 3.91[3.27, 4.50]$ $(range : 0.00 - 6.00)$ pages in the COMPLEX condition (see *Supplementary Information* for further details). While rates of continued listening were numerically greater in the SIMPLE condition, this difference was not significant (t(43.40)=1.54; $p = 0.132$).

Table 1 reports the median values for children's net gaze durations and percent net gaze durations to each AOI, by condition. When listening to the SIMPLE speech—compared to when listening to the COMPLEX speech—children showed greater net gaze durations to the ILLUSTRATION (t(38.01)=2.73, $p < .01$), but similar gaze to the DISTRACTOR (t(35.99)=-0.45, $p = 0.652$).[5] Children's percent net gaze durations did not appear to differ significantly between the two conditions for either the ILLUSTRATION (t(34.89)=1.74, $p = 0.091$) or the DISTRACTOR (t(42.29)=-1.51, $p = 0.138$).

Does complexity level interact with age in predicting children's attention?

Table 2 shows the results for models testing the hypothesis that children's attention to the speech will reflect an interaction between our speech complexity manipulation and children's own cognitive/linguistic development (operationalized here as age).

[5] Recall that variability in children's gaze to the distractor was likely truncated by the mechanics of the experiment, which, after the first repetition of the current page, transitioned to the next as soon as one of children's fixations to the DISTRACTOR cleared the 1500*ms* threshhold.

Table 1
Metrics of Attention to Each AOI by Condition

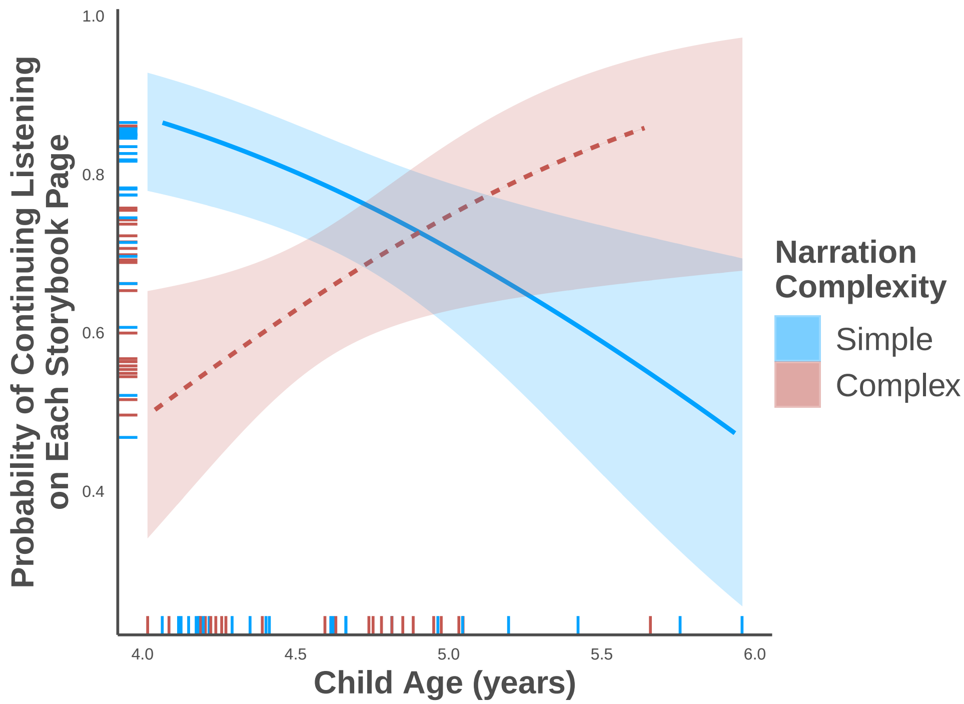
		Simple		Complex	
		Mdn	*95% CI*	*Mdn*	*95% CI*
By *Page*	Net Gaze Duration (*s*)				
	Illustration	14.76	(11.97, 20.83)	12.66	(7.84, 17.72)
	Distractor	5.42	(3.74, 8.28)	5.64	(3.19, 8.80)
	Percent Gaze Duration (%)				
	Illustration	50.55	(40.84, 64.24)	44.38	(26.91, 61.15)
	Distractor	27.93	(17.18, 50.59)	37.05	(21.41, 53.81)
By *Participant*	Total Gaze Duration (*s*)				
	Illustration	97.99	(82.59, 117.58)	80.21	(56.50, 92.72)
	Distractor	41.40	(29.88, 44.68)	39.68	(27.20, 49.47)
	Mean Percent Gaze Duration (%)				
	Illustration	50.22	(45.25, 56.93)	41.77	(34.40, 57.46)
	Distractor	30.54	(23.65, 42.73)	37.02	(31.92, 46.96)

Table 2
Mixed Effects Logit Model of Children's Probability of Continuing Listening

Intercept	4.62*** (2.50, 9.80)
Condition(Complex)	0.47 (0.19, 1.07)
Age	0.56** (0.32, 0.93)
Complex:Age	3.37** (1.42, 9.23)
Observations	276
Subjects	46
Log Likelihood	−151
AIC	314
BIC	335

*p<0.05; **p<0.01; ***p<0.001

Model syntax: `glmer(continued_listening ~ age + condition + age:condition + (1|subject))`



As predicted, there was a significant interaction between condition and age in predicting children's probability of continuing listening (Wald's $\chi^2(1) = 7.2,\ p < .01$; Figure **??**). Specifically, when listening to the Simple speech, older children were less likely to continue listening than younger children (Age $OR = 0.56$ $[0.32, 0.93]$), but the opposite was true for children listening to the Complex speech: older children in the Complex condition were *more* likely than younger children to continue listening on each page ($OR = 3.37$ $[1.42, 9.02]$). This pattern of results is consistent with the idea that the Simple speech represented an appropriate level of complexity for the younger children in our study, whose attention it seemed to elicit and maintain. Older children, though, may have been more likely to have learned all they could from the first repetition of each page, and thus to disattend (gravitate toward the Distractor) when the narration began to loop. That the Complex speech was more likely to maintain the attention of older children suggests that their greater language skills may have made them better able to recognize that there was 'something to learn.'

We saw mixed results with our two measures of children's gaze to the Illustration. As suggested by the descriptive statistics in the previous section, condition was a significant predictor of children's Illustration gaze durations ($beta = -0.48$ $[-0.93, -0.02]$; Wald's $\chi^2(1) = 4.59,\ p < .05$), such that children tended to spend less time looking at the Illustration when listening to the complex speech. For children's gaze *durations*, the interaction between condition and age did not significantly improve model fit ($beta = -0.33$ $[-0.82, 0.16]$; Wald's $\chi^2(1) = 1.85,\ p = 0.173$). However, we saw the inverse for children's *percent* Illustration gaze durations: there, we saw fragile evidence of the predicted interaction between condition and age ($beta = 9.02$ $[0.21, 17.83]$; Wald's $\chi^2(1) = 4.2,\ p < .05$), but no effect of condition ($beta = -7.25$ $[-15.63, 1.12]$; Wald's $\chi^2(1) = 2.98,\ p = 0.084$). The interaction between

Table 3
Mixed Effects Logit Models Predicting Listening Comprehension Accuracy from Child Attention

	Model 1	*Model 2*	*Model 3*
Intercept	2.98* (0.96, 9.99)	3.57* (0.97, 14.70)	3.00* (0.98, 9.93)
Age (mos)	2.09*** (1.51, 3.00)	2.42*** (1.53, 4.23)	2.06*** (1.47, 2.97)
Cond(COMPLEX)	0.87 (0.48, 1.59)	0.82 (0.34, 1.95)	0.85 (0.46, 1.57)
Continued Listening	1.59** (1.17, 2.18)		
ILLUSTRATION Gaze (*s*)		1.61* (1.04, 2.62)	
ILLUSTRATION Gaze (%)			1.48** (1.10, 2.02)
Observations	276	276	276
Log Likelihood	-144	-142	-146
AIC	299	297	301
BIC	317	319	319

*p<0.05; **p<0.01; ***p<0.001

Model syntax: `glmer(correct ~ standardized_attention_index + condition + age + (1|subject) + (1|question))`

condition and age was in the expected direction: relative to younger children, older children directed greater percentages of their gaze to the ILLUSTRATION specifically when listening to the COMPLEX speech (see Supplemental Information for full details).

Are Children's Learning Outcomes and Patterns of Attention Related?

If children's attention to the speech was driven at least in part by their ongoing sense that they were learning from it, we should see a correspondence *across conditions* between measures of individual children's attention and their learning outcomes. Our final analyses test this prediction.

Did children who attended more understand the story better? We fit separate mixed effects logit models to children's listening comprehension test trial accuracy for each index of children's attention to the speech over the course of the story (the *proportion* of storybook pages on which a child continued listening, their *total* ILLUSTRATION gaze duration across pages, and their *mean* ILLUSTRATION percent gaze duration across pages).

In both conditions, children who paid more attention to the narration—according to our three measures of child attention—showed greater understanding and recollection of the story's plot and characters at test. Controlling for condition and age, the proportion of pages on which a child continued listening (Table 3, Model 1) was significantly related to their listening comprehension scores ($OR = 1.59$ $[1.17, 2.18]$, Wald's $\chi^2(1) = 8.71,\ p < .01$). The same was true of children's total ILLUSTRATION gaze duration ($OR = 1.61$ $[1.04, 2.62]$, Wald's $\chi^2(1) = 4.35,\ p < .05$; Table 3, Model 2), and children's average percent ILLUSTRATION gaze duration ($OR = 1.48$ [1.10, 2.02], Wald's $\chi^2(1) = 6.67,\ p < .01$; Table 3, Model 3).

Table 4

Mixed Effects Logit Models Predicting Unfamiliar Word Generalization from Child Attention

	Model 1	*Model 2*	*Model 3*
(Intercept)	0.67 (0.28, 1.49)	0.64 (0.27, 1.38)	0.68 (0.28, 1.54)
Age	0.80 (0.50, 1.28)	0.95 (0.61, 1.47)	0.85 (0.54, 1.33)
Continued Listening	1.56^{*} (1.03, 2.43)		
ILLUSTRATION Gaze (s)		1.15 (0.84, 1.57)	
ILLUSTRATION Gaze (%)			1.59^{**} (1.14, 2.27)
Observations	132	132	132
Log Likelihood	-83.7	-85.5	-82.2
AIC	175.4	179.0	172.3
BIC	186.9	190.6	183.8

*p<0.05; **p<0.01; ***p<0.001

Model syntax:
```
glmer(correct ~ standardized_attention_index + condition + age + (1|subject)
+ (1|question))
```

Did children who attended more learn the hard words better? Finally, we asked whether our three child attention metrics were positively related to children's generalization of the unfamiliar words tested in the final block of test trials. We fit mixed effects logit models to children's unfamiliar word generalization performance in the COMPLEX condition, where the unfamiliar words that we tested were actually used in the narration. Holding child age constant, two of the three metrics of child attention were significantly related to unfamiliar word generalization test accuracy: the proportion of trials on which children continued listening ($OR = 1.56$ [1.03, 2.43], Wald's $\chi^2(1) = 4.21,\ p < .05$; Table 4, Model 1), and children's mean percent gaze duration to the ILLUSTRATION ($OR = 1.59$ [1.14, 2.27], Wald's $\chi^2(1) = 7.08,\ p < .01$; Table 4, Model 3). The odds ratios from these models suggest that a standard deviation in either attention metric was associated with more or less a one-and-a-half times increase in the child's probability of correctly generalizing the unfamiliar target word at test. Children's total gaze duration to the ILLUSTRATION did not emerge as a significant predictor of their word generalization accuracy ($OR = 1.15$ [0.84, 1.57], Wald's $\chi^2(1) = 0.79,\ p = 0.375$; Table 4, Model 2). These results are consistent with the idea that children's attention to the speech was reflective of their sense that they were or could be learning something. The differences across variables suggest that the overall time that children spent looking at the ILLUSTRATION may not have been as good a cue to their learning from the speech as their tendency to listen more to each page and to *prefer* the ILLUSTRATION over the DISTRACTOR or other regions of the screen.

General Discussion

Here, we sought evidence for the foundational idea that children's attention to different sources of information reflects the degree to which they support their learning. Inspired by the real-life context of storybook-reading, we tested this idea by manipulating both the complexity of the language that children heard (by varying the estimated ages of acquisition of the words used in alternative story narrations) and the capacities of the learners themselves (by testing children across a two-year age range spanning significant growth in the vocabulary and language knowledge). Systematic differences in child attention to the SIMPLE versus COMPLEX speech suggest that our experimental manipulation of speech complexity was effective, and that our novel method left children free to direct their attention between a speech stream offering new opportunities for learning and an alluring distractor.

The strongest support for our hypothesis that children's attention is partially driven by their sense of learning came from the interaction between speech complexity and age in predicting children's probability of continuing listening on each page of the story. Instead of an objective level of speech complexity that elicited continued interest equally across all children, children's probability of continuing listening seemed to depend on what Gerken and colleagues (2011) termed the narration's *subjective complexity.* That is, children's desire to hear further repetitions of the same page depended on the size of the gap between their current linguistic competence and the difficulty of the words used to tell the story. When the gap was small, children wanted to continue listening. When the gap was greater, children tuned out. That older children were less likely than younger children to continue listening in the SIMPLE condition (where there was little for them to learn) is especially critical evidence for suggesting that children's attention reflected their sense of learning, and for ruling out that children merely paid more attention with age. We also saw concordant effects when looking at how children distributed their visual attention across the display in the two conditions. There, the interaction between complexity and child age suggested that, as children matured, they devoted a greater percentage of their overall gaze to the story-relevant ILLUSTRATION only in the COMPLEX condition, consistent with greater attention to speech as it "becomes easier."

There was one unexpected finding: children overall seemed to prefer the SIMPLE speech. However, this finding may be less surprising on closer inspection. First, regardless of whether the language itself offered new material for learning—that is, whether the words were all already in the child's vocabulary—the speech was conveying a story, with content for children to learn. Second, the SIMPLE speech in particular had led to robust learning in same-age children recruited from the same area in previous pilot data, a result we also replicate here. Thus, we knew in advance of our study that the SIMPLE speech at least was appropriate for our full sample, and would support learning of the plot knowledge we tested.

Our study additionally enabled us to directly test the link between children's selective, learning-driven attention and their learning *outcomes.* Previous infant research measuring attention based on complexity or learnability has necessarily employed highly simplified stimuli, with limited potential for assessing learning (Gerken et al., 2011; Kidd et al., 2012, 2014). What's more, relative to studies of infant language development, the idea that low-level processes of attention to spoken language might continue to mediate language development into the preschool years has received little attention (Houston & Bergeson, 2014). Yet across conditions, we found that individual children's self-directed attention to the speech—measured in terms of children's probability of continued listening, their gaze duration to the storybook illustration, and their percent gaze duration to the illustration—was positively related to their plot knowledge, and to their generalization of unfamiliar words encountered in the Complex condition. Our results thus offer a novel contribution to these previous literatures, revealing that individual children's self-directed attention to the speech in our study reflected the fit between their cognitive-linguistic knowledge and the complexity of the speech. Controlling for condition and age, metrics of this attention correlated with both learning outcomes we tested.

Finally, we were particularly interested in children's sensitivity to naturalistic speech complexity as a means of explaining why certain sources of language input have proven to be more useful for children's learning than others. As in other domains where, for example, children track the past accuracy of informants and use it to select who they want as a teacher (Pasquini, Corriveau, Koenig, & Harris, 2007), our results indicate that children may track the relative difficulty of processing and encoding different sources of linguistic information, and preferentially attend to those sources where they sense their learning is most efficient. Future studies will directly test the idea that independent measures of children's level of linguistic knowledge may predict how they allocate attention to language inputs of different levels of complexity in their environment (e.g., overheard speech or news broadcasts), and whether children are able to actively select the best information sources to enhance their own learning.

References

:::

A Language and Environment for Statistical Computing (Vol. 2, pp. https://www.R–project.org). (2020). (Vol. 2). R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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

Bates, Douglas, Mächler, 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

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Caron, R. F., & Caron, A. J. (1969). Degree of stimulus complexity and habituation of visual fixation in infants. *Psychonomic Science*, *14*(2), 78–79.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Gerken, L. A., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid "labouring in vain" by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–979. https://doi.org/10.1111/j.1467-7687.2011.01046.x

Houston, D. M., & Bergeson, T. R. (2014). Hearing versus listening: Attention to speech and its role in language acquisition in deaf infants with cochlear implants. *Lingua*, *139*, 10–25. https://doi.org/10.1016/j.lingua.2013.08.001

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kuperman, V., Stadthagen-Gonzales, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods*, *44*(4), 978–990.

Martin, R. M. (1975). Effects of familiar and complex stimuli on infant attention. *Developmental Psychology*, *11*(2), 178.

Mayer, M. (1969). *Frog, where are you?* Dial Press.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216.

Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1122—1133. https://doi.org/10.1037/a0023101

Thomas, H. (1965). Visual-fixation responses of infants to stimuli of varying complexity. *Child Development*, 629–638.

Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. Retrieved from https://books.google.se/books?id=RxjjUefze%7B/_%7DoC