

Preschoolers selectively attend to speech that they can learn more from

Ruthe Foushee^{1,2,✉}, Mahesh Srinivasan², and Fei Xu²

¹Department of Psychology, New School for Social Research, 80 5th Avenue, New York, NY 10011

²Department of Psychology, University of California, Berkeley, 2121 Berkeley Way West, Berkeley, CA 95720

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 ability to generalize 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 narration complexity and age in predicting children’s probability of continuing listening on each page, and the proportion of their visual attention that they devoted to the story illustration, over the animated distractor. 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.

Correspondence: ruthe@newschool.edu

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

Publication Citation: Foushee, R., Srinivasan, M., & Xu, F. (2025). Preschoolers selectively attend to speech that they can learn more from. *Developmental Science*.

ACKNOWLEDGMENTS

We thank the children and their families for participating, as well as the teachers at the Harold E. Jones Child Study Center, University Village Albany Child Development Center, Haste Street Child Development Center, Kensington Elementary School, and The Berkeley School; Alison Fong, Jacqueline Nguyen, Jon Wehry, Harmonie Strohl, Gwyneth Heuser, Grace Horton, and Luvy Vanegas Grimaud for help with stimuli and data collection; Mike Frank, Susanne Gahl, Terry Regier, and Azzurra Ruggeri for comments on analysis; the members of the Berkeley Early Learning and Language and Cognitive Development Labs at UC Berkeley—including Monica Ellwood-Lowe and Ariel Starr—for feedback on drafts; and Chef Miko for editing. Earlier versions of this work were presented at the 45th Annual Boston University Conference on Language Development, the Biennial Meeting of the Society for Research in Child Development, 2021, and the Annual Meeting of the Jean Piaget Society, 2022. This work was supported by NSF GRFP DGE-1752814 to RF and NSF SMA-1640816 to FX. This article was written using the papaja library in RStudio (Aust & Barth, 2020).

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 (e.g., Bruner, 1961; Vygotsky et al., 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 (Foushee et al., 2023; Gureckis & Markant, 2012).

Background

Results from infant studies are consistent with the idea that children’s attention to a stimulus is driven by their sense of learning. In one body of work, 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 (e.g., Caron & Caron, 1969; Kidd et al., 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 conditional probability. 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 visual attention in response. The authors dubbed the pattern they observed the “Goldilocks effect”: infants’ probability of looking away 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. Importantly, this ‘U-shaped’ relation between stimulus complexity and infants’ probability of looking away was evident not just at the group level, but in individual infants’ gaze behavior, at different ranges along the complexity continuum. This is what we would expect if infants’ attention was driven by their sense of learning, because different complexity ranges will be relevant and appropriate for different infants. However, while attending to intermediate complexity is understood as a domain-general learning mechanism, studies showing complexity-based attentional preferences are typically not designed to directly demonstrate the learning payoff 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 a step closer toward linking selective attention and learning show how individuals’ attention shifts with experience. For example, Forest et al. (2022) show how adults attend to increasingly complex patterns as they gain more experience with sequential visual stimuli. Poli et al. (2020) link infants’ attention to their learning progress: In their study, 8-month-olds watched individually-cued target shapes reappear at different locations on a screen. Each shape had a most-likely target location, making some trials more informative than others for ultimately predicting where a given shape was going to show up. Infants’ gaze in this paradigm showed the established relationship between complexity and attention: infants were least likely to look away for trials of intermediate predictability. However, *learning progress*, or how informative a given trial was, proved an even stronger predictor of infant looking times: infants were least likely to look away when information gain was highest. Infants’ *actual* learning progress was evident in their looking times across trials: infants became faster and faster at directing their gaze toward predictable targets, consistent with having developed an efficient model of the statistical environment. Together, these studies confirm that learners’ attention is informed by relative complexity. Further, they show that relative complexity is a moving target, informed by what learners have already seen (and it is notably almost always “seen,” as most studies have shown these effects in the visual domain.) These studies also employ specific notions of complexity and learning, compatible with their stripped-down visual-event stimuli: ‘complexity’ means lower conditional probability; ‘learning’ means being able to efficiently predict the next event.

Our study aimed to test this theoretically domain-general learning mechanism in natural language—a messy, multiplex domain with real-world consequences. Using meaningful linguistic stimuli enables us to probe relations among stimulus complexity, selective attention, and learning not just for sequential statistical dependencies, but for the higher-order

sense-making involved in language comprehension at older ages. Two follow-ups on Kidd et al.’s original study are especially relevant, given this focus. The first shows the Goldilocks effect in *auditory* attention, measuring infants’ probability of looking away from a display in response to tone sequences of varying predictability (Kidd et al., 2014). The second tests children 3 to 6 years of age, finding the same relation between the predictability of visual event sequences and children’s probability of looking away (Cubit et al., 2021). To our knowledge, only one other study has used *linguistic* stimuli to home in on the hypothesis that infants attend more to information that they would be more likely to learn from: In one of two experiments, Gerken and colleagues (2011) exposed half of 17-month-old infants to artificial language stimuli that had supported grammar learning and generalization in prior samples of same-age infants (making it *subjectively learnable* by infants of this age). The other half of infants heard artificial language stimuli that had *failed* to support learning and generalization of the corresponding grammar in previous samples (making it *subjectively unlearnable*). Infants took longer to habituate to the subjectively learnable stimuli, leading the authors to propose that infants implicitly monitor their learning rate from a particular information source, and disattend when it falls below some threshold of efficiency.

Notably, while learning is implicated as the underlying motivation for children’s attention, previous studies have not directly tested children’s learning from the same stimuli to which attention is measured (Poli and colleagues’ implicit tracking of learning progress is a notable exception). Existing studies 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 a product of the interaction between stimulus complexity and the relative competence of the learner; that is, the same “objective” level of complexity will be less subjectively learnable for a more novice learner, and more subjectively learnable for a more advanced learner. Thus, any study that only varies stimulus complexity cannot be sure that attentional preferences among children at the same level of development are a result of *learnability*, over alternate dimensions of the stimulus.

The Present Study

The current study addresses these gaps using a novel paradigm to test 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 that employ highly simplified visual or auditory stimuli (Cubit et al., 2021; Forest et al., 2022; Kidd et al., 2012, 2014; Poli et al., 2020), 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

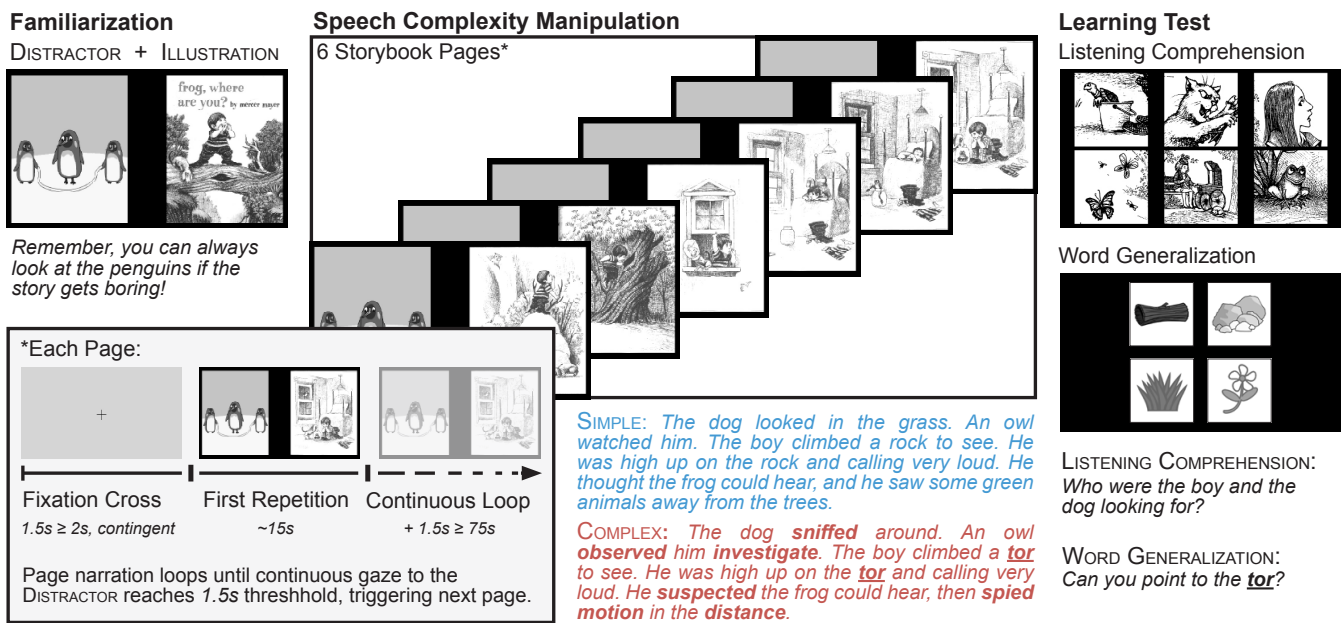


Fig. 1. Schematic of Experimental Eyetracking Procedure Manipulating the Speech Complexity of a Narrated Storybook and Measuring Child Attention and Learning.

story contained many words that were likely to be unfamiliar. During the story narration, we measured children’s attention to the speech via their gaze to story-relevant visual stimuli. After the story narration, we measured children’s learning from the speech via explicit tests of their listening comprehension and partial word knowledge.

What do we expect to see? In thinking through 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 of course expect that the amount a child learns from a stimulus will be related to the amount that the child attended to it.¹ 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 whether a child’s attention allocation is itself determined by the interaction between stimulus complexity and the child’s own competence, as this determines the child’s capacity to learn from the stimulus. In our study, we use children’s age as a proxy for their relative cognitive development and linguistic competence.² We expect that there will be a larger gap between the COMPLEX speech

¹While seemingly obvious, we note that this remains untested in previous work.

²In a previous sample of same-age children recruited from the same venues as our current participants, children’s age was highly correlated with their raw scores on the Peabody Picture Vocabulary Test ($r = .74, p < .001$; Dunn & Dunn, 1981), a more narrowly relevant measure of competence. Age was also significantly correlated ($r = .39, p < .001$) with children’s learning of the same set of rare words embedded in the COMPLEX speech condition of the current study (see *Materials and Methods*). Both of these facts suggest that age is a reasonable index to use in predicting the relative ‘learnability’ or ‘subjective complexity’ of linguistic stimuli whose objective complexity we have manipulated.

and the language that the younger children in our sample know (making it difficult for them to learn from), and we expect this gap to narrow as children age (making the COMPLEX speech more learnable 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 in our sample than from the older children.

To test these predictions, we tracked children’s visual attention to a display while listening to either the SIMPLE or COMPLEX narration of a textless storybook (Mayer, 1969)—distinguished by the ages of acquisition of the words the narrations used (Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012). We directly tested children’s learning outcomes from the story narration, including their comprehension of its plot and their ability to understand and generalize rare words employed in the story. On each page of the story, the storybook ILLUSTRATION competed for children’s visual attention with the DISTRACTOR (a continuous animation of three penguins double-dutching; see Figure 1). Given the presence of this dynamic animation, 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 inadvertently revealed by a child’s gaze; Cooper, 1974; Salverda and Altmann, 2011). Indeed, when the same split-screen display was presented *sans* narration during piloting, children

looked almost exclusively at the DISTRACTOR, rather than at the ILLUSTRATION. Thus, we expected that children would be lured by the DISTRACTOR as soon as they were no longer listening to the story. This basic design originated from even earlier pilot studies in which children heard the story narration while watching a screen with *only* the storybook illustrations. Pilot eyetracking data revealed a large amount of ‘lost’ gaze data—children’s gaze falling off-screen, in the white space around the image, or otherwise proving untrackable—which seemed to coincide with behavioral indices of children’s loss of interest in the speech (fidgeting, humming, etc.). In introducing the DISTRACTOR, we hoped to capture rather than lose those ambiguous attentional data, and to refine the interpretation of children’s non-ILLUSTRATION gaze as an inverse signal of their attention to the speech.

Inspired by the trial structure of gaze-contingent infant paradigms (Colombo & Mitchell, 2009), the duration of each storybook page was contingent on children’s allocation of visual attention. While 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), children who were consistently drawn in by the DISTRACTOR moved through the story quickly and heard the narration for each page only once. This design met the twin goals of making sure that every child: (a) heard the entirety of the story content (so that variation in learning outcomes could not come from variation in who had even heard the critical information); (b) provided attentional data for *all* pages (i.e., even if the child would otherwise have disattended from the entire story—and experiment—at an earlier point).

We quantify children’s attention to the story by measuring whether children continued listening to further, optional repetitions of the narration for each page, as well as how much they looked to the ILLUSTRATION (which is only made salient by the narration), over the DISTRACTOR (which is salient otherwise).

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 capacity to generalize rare target words embedded in the speech to novel referents. These variables enabled us to answer the following specific research questions: (a) is preschool-aged children’s attention to naturalistic speech responsive to its complexity? (b) does children’s age—as a proxy for their level of language and cognitive development—interact with the experimentally manipulated complexity of the speech in predicting their attention (consistent with children’s attention to spoken language being sensitive to how much they can learn from it)? and (c) within each condition (SIMPLE/COMPLEX), is children’s attention to the speech correlated with their learning outcomes?

Materials and Methods

Full documentation of our procedures, including study scripts and stimuli, is at https://osf.io/zsjfb/?view_only=024c8e83e56a4fff95e5d5ae840035c2. Session videos are available on databrary.org (linked in the study repository), for viewing by registered users at the access level permitted by each participating family. Deidentified data and analysis files, along with a reproducible version of this manuscript, are at <https://github.com/foushee/storybook-habitation>.

1.1 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. Based on demographic information provided by 83% ($N = 38$) of caregivers, 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; the median household income in Berkeley, CA was \$91,259 in 2020; U.S. Census Bureau, 2020). 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 The Institute of Human Development at the University of California, Berkeley. They were tested in a quiet area of their school or in lab, and received a sticker and/or certificate and small toy for their participation. The study was approved by the University of California, Berkeley Committee for the Protection of Human Subjects.

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

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.

1.1.1 Vocabulary Questionnaire 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, administered via Qualtrics

(<https://www.qualtrics.com>). For every content word used in either condition of the study, caregivers indicated whether or not their child would “understand the word if [the caregiver] said it out loud.” Caregivers typically filled out this measure while their child was participating in the study, along with a demographic survey and language environment questionnaire. Caregivers of children tested in preschool received a link to the questionnaires via e-mail. This survey confirmed that the rare target words embedded in the COMPLEX condition were indeed novel to children: 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). On the other hand, caregivers reported that their children would understand *all* of the words used in the SIMPLE condition.

1.2 Procedure

1.2.1 Familiarization Children sat before a laptop connected to an SMI RED-n eyetracker, wearing child-sized over-ear headphones. After a brief five-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.

1.2.2 Storybook Narration The same female voice narrated a boy and 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 gated the onset of each new page (see inset of Figure 1).

Speech Complexity Manipulation Depending on the condition to which they were assigned, children heard the story narrated at either the SIMPLE or COMPLEX level (center column of 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-Gonzalez, & Brysbaert,

2012).³ 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 (~ 15s), after which the same audio continued to loop for up to five additional repetitions (~ 75s), separated by a brief pause (~ 500ms). After the first obligatory narration, 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 1.67s and 6.70s ($M = 2.97s$ [2.64, 3.31]).

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 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!’).

1.2.3 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. Within each block, each test trial tested knowledge from a different content page in 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, mirroring the order in which the relevant information was introduced within the story.

Listening Comprehension In the first test block, *listening comprehension* trials tested children’s knowledge of story events or characters (equivalently presented across the SIM-

³ Across pages, narrations were matched for syllable count (*range* = 50 – 61 syllables, $M = 54.92$ [53.33, 56.59]; $t(5.00) = 0.15$, $p = 0.889$ paired by page), speech rate (*range* = 3.42 – 3.99 syllables/second, $M = 3.67$ [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, $M = 0.90$ [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 once each, while the SIMPLE condition used a more accessible alternative. Comparisons between the language of the two conditions can be found in the online repository for this study.

PLE and COMPLEX conditions, just using easier or more difficult synonyms). 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 2×3 grid of black-and-white images (all drawn 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 to identify appropriate illustrations for the unfamiliar target words embedded (exclusively) in the COMPLEX narration. For the children in the COMPLEX condition, this meant generalizing the unfamiliar target words that they had heard 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).⁴ 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 2×2 grid. Competitor images were selected to be compatible with the syntax of the test question (e.g., depicting other actions with thematic patients as options for ‘ogling’). The correct response for all questions was normed via a sample of undergraduates exposed to the same story narration ($N = 19$).

1.3 Variable Coding and Predictions

1.3.1 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 the (2) duration and (3) distribution of children’s visual attention to our predefined Areas of Interest (AOIs; the ILLUSTRATION and DISTRACTOR).

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 had ended, plus the 1500ms threshold 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 meaningfully analyze children’s voluntary exposure to the speech, in spite of the challenges presented by children’s raw voluntary listening durations (namely, zero-inflation—many children moved on to the next page shortly after the first repetition—and a long tail; see *Supplementary Information*).

We use this recoded variable for analyses at both the page-

⁴The test trial for ‘absconding’ also used a different tense (present progressive) from the story, where both uses were in the simple past: “. . . the frog absconded from the jar. He absconded to find his mom and dad.”

and subject- levels. At the page level, continued listening is a binary variable. Across all pages, children moved on immediately 29% of the time, and listened to all five additional repetitions less than 2% of the time (just 5 individual storybook pages). At the subject level, we analyze continued listening in terms of the *proportion* of pages on which a child continued listening ($range = 0 - 1$ $M = 0.71$ [0.64, 0.78]). Approximately half of all children ($N = 24$; 52%) continued listening on at least 5 out of the 6 total pages of the story. We take this measure to reflect a child’s ongoing attention to the speech, or their sustained appetite for hearing more of it. Conversely, we can think of preschoolers’ probability of ‘moving on’ from a storybook page in the present paradigm as analogous to infants’ probability of looking away in previous research, interpreted minimally as a loss of interest in the stimulus (Rankin et al., 2009), and more richly as an active decision to reallocate cognitive resources away from it (Kidd & Hayden, 2015)).

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 children’s 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*)⁵ 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 page level, each child contributed twelve net gaze durations: one value for each of the two AOIs, on each of the six storybook pages (ILLUSTRATION: $range = 0 - 62.78s$, $M = 15.42$ [14.38, 16.44]; DISTRACTOR: $range = 0 - 24.51s$, $M = 6.59$ [6.12, 7.04]). When analyzing net gaze durations at the subject level, we sum 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 ($range = 27.38 - 148.47s$, $M = 90.09$ [82.30, 97.92]; *total DISTRACTOR gaze duration*: $range = 14.11 - 73.11s$, $M = 39.54$ [35.79, 43.47]).

Proportion Gaze Duration A child’s *proportion gaze duration* for a given AOI represents their gaze to that AOI as a proportion of their gaze across the entire display.⁶ This measure narrows in on the relative *share* of children’s visual attention devoted to each AOI (ILLUSTRATION: $range = 0 - 0.86$, $M = 0.47$ [0.45, 0.50]; DISTRACTOR: $range = 0 - 0.94$, $M = 0.35$ [0.33, 0.38]), irrespective of overall duration. As

⁵While we report descriptive statistics for gaze durations in *seconds* for readability (Table 1), we use log-transformed *ms* values in our page-level statistical models.

⁶The majority of children’s gaze to the display— $M = 83%$ [81, 85] across pages—was typically captured by one of our two AOIs. In other words, children mostly looked at either the ILLUSTRATION or the DISTRACTOR, and spent a small minority of their looking time in the white space surrounding the two images, e.g., when switching between them.

a page-level index of attention to the story, we analyze children's proportion gaze durations to the ILLUSTRATION, specifically. We average this value across pages for a subject-level metric (*mean ILLUSTRATION proportion gaze duration: range = 0.18 – 0.71, M = 0.47 [0.43, 0.51]; mean DISTRACTOR proportion gaze duration: range = 0.07 – 0.77, M = 0.35 [0.31, 0.40]*).

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.

We note that the attention metrics that we have included are likely to be intercorrelated: When children 'continue listening,' they have a greater window in which to gaze at the ILLUSTRATION, making longer ILLUSTRATION gaze durations more likely. Looking more to the ILLUSTRATION will likely correspond with looking less to the DISTRACTOR, resulting in greater ILLUSTRATION proportion gaze durations. And devoting more attention to the DISTRACTOR, relative to the ILLUSTRATION, might mean that the child is already looking at the DISTRACTOR when the first narration ends, causing them to 'move on' as soon as possible, rather than continue listening. Despite their likely intercorrelations, we nevertheless chose to include all three metrics because they offered distinct analytic and interpretive advantages. First, by splitting children's listening times into the binary categories of 'moved on' vs. 'continued listening,' we maximize our ability to detect differences in children's attention patterns on the basis of condition and age, and provide an analogy to the look-away probability measure employed in infant research. Second, insofar as they pull apart, our two metrics of children's visual attention to the ILLUSTRATION may allow us to infer whether learning is driven more by increased looking toward relevant stimuli (greater ILLUSTRATION gaze durations) vs. decreased looking toward irrelevant stimuli (greater ILLUSTRATION gaze proportions).

1.3.2 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%]*).

Unfamiliar Word Generalization Children's responses on the trials testing the unfamiliar words in the COMPLEX narration were likewise coded as correct (1) or incorrect (0). Children showed predictably variable performance on this measure. However, while we expect the variation in performance among children in the COMPLEX condition (*range = 0 – 83% correct, M = 39% [31, 48]*) to be potentially meaningful, we expect the variation in the SIMPLE condition—where children never heard the target words—to merely reflect chance (*range = 0 – 83% correct, M = 31% [24, 40]*). Thus, we analyze unfamiliar word generalization accuracy only in the COMPLEX condition, where children actually had the opportunity to learn something about the words from within the experiment.

If children's attention is at least partly sustained by their sense that they are or could be learning something, we expect that these two measures of individual children's learning outcomes—evidence that they indeed learned from the speech—will correlate with measures of the same children's attention to the speech, described in the preceding section.

1.4 Analysis

In addition to reporting descriptive statistics regarding our variables of interest, we conduct two primary varieties of analyses: (1) analyses testing the link between speech complexity and child attention, and (2) analyses testing the link between child attention and learning outcomes.

In the first set of analyses, we fit separate models to the page-by-page data for each 'positive'—as in, predicted to be associated with learning—attention metric, according to its distribution: We use the `lme4` package (Bates et al., 2015) in R (v4.1.2; R Core Team, 2021) to fit mixed effects logit models to children's binary continued-listening codes,⁷ and linear mixed effects models to children's continuous ILLUSTRATION net gaze durations.⁸ We use the `mgcv` package (Wood, 2011) to fit a generalized additive model to children's ILLUSTRATION proportion gaze durations, assuming a beta distribution.⁹ In each of these models predicting page-by-page

⁷Model syntax: `glmer(continued_listening ~ age + condition + age:condition + (1|subject) + (1|page), family="binomial"(link="logit"))`

⁸Model syntax: `lmer(gaze_duration ~ age + condition + age:condition + (1|subject) + (1|page))`

⁹Model syntax: `gam(gaze_proportion ~ age + condition + age:condition + s(subject, bs="re") + s(page,`

Table 1. Metrics of Attention to Each AOI by Condition

		SIMPLE		COMPLEX	
		<i>Mdn</i>	<i>95% CI</i>	<i>Mdn</i>	<i>95% CI</i>
By Page	Net Gaze Duration (<i>s</i>)				
	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)
	Proportion Gaze Duration				
	ILLUSTRATION	0.51	(0.41, 0.64)	0.44	(0.27, 0.61)
	DISTRACTOR	0.28	(0.17, 0.51)	0.37	(0.21, 0.54)
By Participant	Total Gaze Duration (<i>s</i>)				
	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 Proportion Gaze Duration				
	ILLUSTRATION	0.50	(0.45, 0.57)	0.42	(0.34, 0.57)
	DISTRACTOR	0.31	(0.24, 0.43)	0.37	(0.32, 0.47)

attention, 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 (or random effect smooths, for the generalized additive model) for child and page. In cases where a model of this structure fails to converge, we refit the model after dropping the random effect with the lowest variance (Barr et al., 2013).

In the second set of analyses, we use mixed effects logit models to predict children’s trial-by-trial test question accuracy. We fit models to predict accuracy from each subject-level attention variable (*continued listening proportion*, *total ILLUSTRATION net gaze duration*, *mean ILLUSTRATION proportion gaze duration*) separately, controlling for condition (*listening comprehension* models only) and age (*listening comprehension* and *unfamiliar word generalization* models). We standardize the attention predictor variables ($M = 0$; $SD = 1$) to enable us to compare effect sizes across them. Models include random intercepts for child and question.¹⁰

We rely on estimates and 95% bootstrapped confidence intervals of model coefficients or odds ratios (ORs) to interpret the impact of different predictors on the dependent variable. We assess the significance of individual predictors by comparing nested models with and without the relevant predictor (using the `anova` function in R; R Core Development Team, 2021).

`bs="re"), family=betar(link="logit"))`
¹⁰Model syntax: `glmer(question_correct ~ standardized_attention_index + condition + age + (1|subject) + (1|question), family="binomial"(link="logit"))`

Results

2.1 Is Children’s Attention Responsive to Spoken Language Complexity?

2.1.1 Do Children Differentially Attend to the Simple vs. Complex Speech? Children continued listening on an average of $M = 4.62$ [4.00, 5.21] ($range = 1 - 6$) pages in the SIMPLE condition, and $M = 3.91$ [3.27, 4.50] ($range = 0 - 6$) 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 proportion 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 total ILLUSTRATION net gaze durations ($t(38.01)=2.73$, $p < .01$), but similar total DISTRACTOR net gaze durations ($t(35.99)=-0.45$, $p = 0.652$).¹¹ Children’s mean proportion 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 ($t(42.29)=-1.51$, $p = 0.138$).

2.1.2 Does Complexity Level Interact with Age in Predicting Children’s Attention?

Table 2 shows the results for a model 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 via children’s age. As pre-

¹¹Recall 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 page as soon as one of children’s fixations to the DISTRACTOR crossed the 1500ms threshold.

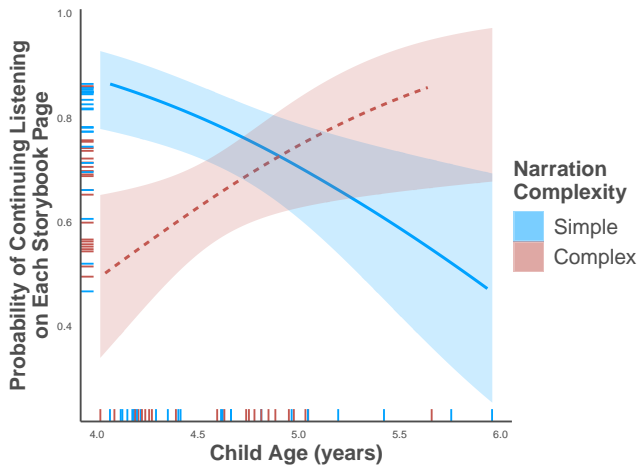


Fig. 2. Model-estimated Probabilities of a Child Listening to Further Optional Repetitions of Each Storybook Page Narration. Child age (x-axis) was associated with a significant decrease in the probability of continuing listening in the Simple condition, but a significant increase in the probability of continuing listening in the Complex condition. Shaded region indicates SE; hatches represent individual children, colored by condition.

dicted, there was a significant interaction between condition and age in predicting children's probability of continuing listening (Wald's $\chi^2(1) = 7.20, p < .01$; Figure 2). 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: in the COMPLEX condition, older children were *more* likely than younger children to continue listening on each page (COMPLEX:Age $OR = 3.37 [1.42, 9.02]$). 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 was more likely to elicit and maintain. Older children may have been more likely to have learned all they could from the first repetition of each page, and thus to disattend (grav-

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, 9.02)
Observations	276
Subjects	46
Log Likelihood	-151
AIC	314
BIC	335

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

itate toward the DISTRACTOR) when the narration began to loop. The fact that the COMPLEX speech was more likely to maintain the attention of older children is consistent with the idea that their greater cognitive and linguistic 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 attention to the ILLUSTRATION. For our absolute measure (children's net ILLUSTRATION gaze durations), only condition was a significant predictor, such that children looked less to the ILLUSTRATION in the COMPLEX condition ($\beta = -0.29 [-0.51, -0.06]$; Wald's $\chi^2(1) = 6.48, p < .05$). Neither age ($\beta = -0.05 [-0.19, 0.09]$; Wald's $\chi^2(1) = 0.14, p = 0.712$), nor the interaction between condition and age ($\beta = 0.08 [-0.15, 0.32]$; Wald's $\chi^2(1) = 0.48, p = 0.487$) significantly predicted gaze durations. Turning to our *relative* measure of visual attention to the ILLUSTRATION: age, condition, and their interaction were significant predictors of children's proportion ILLUSTRATION gaze durations. Children directed lesser proportions of their gaze to the ILLUSTRATION when listening to the COMPLEX speech ($OR = 0.56 [0.32, 1.00]$; Wald's $\chi^2(1) = 3.90, p < .05$). The interaction between condition and age was in the expected direction: when listening to the SIMPLE speech, older children tended to direct a lesser proportion of their gaze to the ILLUSTRATION, relative to younger children ($OR = 0.70 [0.49, 1.00]$; Wald's $\chi^2(1) = 3.94, p < .05$). When listening to the COMPLEX speech, older children directed a greater proportion of their gaze to the ILLUSTRATION, relative to younger children ($OR = 2.05 [1.12, 3.74]$; Wald's $\chi^2(1) = 5.44, p < .05$; see *Supplemental Information* for full details).

2.2 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 learning outcomes and measures of individual children's attention. Our final analyses test this prediction separately for each learning outcome and attention metric.

As anticipated, the positive subject-level attention metrics on which these analyses rely were significantly intercorrelated (*continued-listening proportion* and *total ILLUSTRATION gaze duration*: Spearman's $r(44) = 0.60, p < .001$; *total ILLUSTRATION gaze duration* and *mean proportion ILLUSTRATION gaze duration*: Spearman's $r(44) = 0.80, p < .001$; *mean proportion ILLUSTRATION gaze duration* and *continued-listening proportion*: Spearman's $r(44) = 0.57, p < .001$). We interpret differences in how results for each attention metric conform to our theoretical predictions in the General Discussion.

2.2.1 Did children who attended more understand the story better? Children tended to perform well on trials testing their listening comprehension (SIMPLE: *range* = 0 – 100% accuracy, $M = 70\% [59, 80]$; COMPLEX: *range* = 17 – 100%,

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

	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>
(Intercept)	2.98* (0.96, 9.99)	3.57* (0.97, 14.70)	3.00* (0.98, 9.93)
Age	2.09*** (1.51, 3.00)	2.42*** (1.53, 4.23)	2.06*** (1.47, 2.97)
Condition(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 Proportion			1.48** (1.10, 2.02)
Observations	276	276	276
Log Likelihood	-144	-142	-146
AIC	299	297	301
BIC	317	319	319

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

$M = 64\%$ [55, 74]). To test the relation between this learning outcome and children’s online attention to the speech, we fit separate mixed effects logit models to their 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 proportion gaze duration).

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 (see Table 3). 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 accuracy ($OR = 1.59$ [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 mean proportion ILLUSTRATION gaze duration ($OR = 1.48$ [1.10, 2.02], Wald’s $\chi^2(1) = 6.67$, $p < .01$; Table 3, Model 3). Odds ratios for these different subject-level attention metrics were similar, such that a standard deviation increase in each was associated with about a one-and-a-half times increase in the child’s probability of answering a test question correctly.

2.2.2 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 three models showed significant relations between child attention and unfamiliar word generalization test accuracy: these were the proportion of trials on which children continued listen-

ing ($OR = 1.56$ [1.03, 2.43], Wald’s $\chi^2(1) = 4.21$, $p < .05$; Table 4, Model 1), and children’s mean proportion 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). Similar to our listening comprehension results, the odds ratios from these models suggest that a standard deviation increase in either attention metric was associated with more or less a one-and-a-half times increase in children’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—though by no means decisive proof of—the idea that children’s attention to the speech was reflective of their implicit 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 children’s online learning from the speech as their tendency 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 the sources of information 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 across a more simple vs. more complex narration of the same story) and the capacities of the learners themselves (by testing children across a two-year age range spanning significant growth in vocabulary and language knowledge; Brown, 1973). Systematic differences in child attention while listening to the SIMPLE versus COMPLEX story narrations suggest that our experimental manipulation of speech complexity was effective, and that our novel method

Table 4. Mixed Effects Logit Models Predicting Unfamiliar Word Generalization from Child Attention

	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>
(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 Proportion			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

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 was at least partially driven by their sense of learning comes from the interaction between speech complexity and age in predicting children’s (a) probability of continuing listening on each page of the story and (b) ILLUSTRATION gaze proportion. Our results suggest that children’s attention depended on the narration’s *subjective* complexity, rather than on an objective level whose effects or attractiveness was preserved across children. 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 from a second repetition of the same page) is especially critical evidence that children’s attention reflected their sense of learning, because it helps rule out the possibility that children merely paid more attention with age. We saw similar effects when looking at how children distributed their visual attention across the display in the two conditions. Relative to younger children, older children devoted a greater proportion of their overall gaze to the story-relevant ILLUSTRATION only in the COMPLEX condition, consistent with the idea that children attend more to speech as it becomes less subjectively complex.

There was one unexpected finding: children overall seemed to attend more to the SIMPLE speech. On further consideration, there are a few reasons why we might have anticipated this result. First, the story was novel, even if the language was not. We had initially expected older children to exhibit significantly less attention to the SIMPLE speech because it would hold little learning opportunity for them. However, regardless of whether the language itself offered new material for learning—that is, for example, whether the

words and their usages were already familiar to children—the speech was conveying a *story* that children did not already know. Thus, even if learning were the primary way to secure children’s attention, we would expect the SIMPLE narration to be attractive to both younger and older children. Second, the U-shaped curve associated with the Goldilocks effect is not symmetric. Infants and young children disattend more decisively to too-complex stimuli than to too-simple stimuli (Cubit et al., 2021; Kidd et al., 2012, 2014), suggesting that children may have been less likely to disattend from the SIMPLE speech because it was ‘too easy’ than from the COMPLEX speech because it was ‘too hard’. Third and finally, our sample skewed young. In previous pilot data, the SIMPLE speech led to robust listening comprehension performance in same-age children recruited from the same area (a result we replicate). This means that we knew in advance of our study that children across our sample age range could at least understand and learn from the SIMPLE speech. In contrast, only older children in previous samples had reliably learned from the COMPLEX speech. Thus, it is likely relevant for explaining the overall SIMPLE speech advantage that our sample skewed toward the younger end of our age range (see hatches along the x-axis in Figure 2).

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). Studies inferring infants’ learning from their gaze behavior have been limited to the visual domain, and to an operationalization of learning as event prediction (Poli et al., 2020). Across conditions, we found that individual children’s self-directed attention to the speech—measured in terms of children’s probability of continued listening, ILLUSTRATION gaze duration and ILLUSTRATION gaze proportion—was positively related to their plot knowledge. Two of these three attention metrics—children’s tendency to continue listening

and their mean ILLUSTRATION gaze proportions—were also positively related to children’s generalization of the unfamiliar target words embedded in the COMPLEX condition. Our results thus offer a novel contribution to previous studies, suggesting that individual children’s self-directed attention to speech reflects the fit between children’s cognitive-linguistic knowledge and the objective complexity of the speech.

Despite their significant intercorrelations, we did not observe all the predicted effects for all the positive attention variables. Most notably, we saw distinct results for the two measures of children’s attention to the ILLUSTRATION. The proportion of children’s overall gaze that was devoted to the ILLUSTRATION was significantly predicted by the interaction of complexity and age, but their absolute gaze duration was not. Likewise, children’s mean ILLUSTRATION gaze proportions predicted their unfamiliar word generalization, but their total ILLUSTRATION gaze durations did not. This pattern lends itself to a few mutually compatible interpretations. First, net gaze duration may just be a noisy measure of attention. There are many routes to high values, including even children ‘zoning out’ (Shepherd & Kidd, 2024). Another route is via many shorter fixations, potentially broken up by looks to the DISTRACTOR that repeatedly fail to meet the threshold for triggering the next page. Relative and absolute measures of children’s gaze to the ILLUSTRATION pull apart in this latter scenario, as *proportion* gaze duration would be comparatively low (given that children would be effectively splitting their time between the ILLUSTRATION and the DISTRACTOR). This points to another, related interpretation of the discrepancy between measures: it may be that in this experimental setup, what is really important for us to track is how much children get distracted, rather than how much they attend, or how much exposure they get to the speech—and proportion gaze duration may be a particularly good metric of a child’s degree of distraction. Lastly, differences across attention variables might derive from our analytic approach. For example, our power simulations anticipated a crossover effect in predicting attention from condition and age. For an interaction where the effect of age on a particular variable is merely attenuated in one condition versus another (rather than reversed), our study is likely underpowered.

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 (e.g., the quantity of simplified child-directed language that children receive reliably correlates with early vocabulary growth, while the quantity of—often more complex—language around them does not; Ramírez-Esparza et al., 2014; Shneidman and Goldin-Meadow, 2012; Shneidman et al., 2013; Weisleder and Fernald, 2013). 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 our study suggests that this is likely the case. 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 et al., 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 their learning is most efficient. Future studies could 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; Foushee et al., 2016), and whether children are able to actively select the best information sources to enhance their own learning. More generally, language is an interesting test domain in part because language-learning is a lifelong endeavor—not only do adults continue to learn new vocabulary in their first language(s), but studying a second language puts them back in the position of being sensitive to the subjective complexity of different sources of language. Future work might also investigate how differing levels of proficiency across a given child’s or adult’s languages predict the complexity of linguistic input that they spontaneously attend to in each.

3.1 Limitations

Our study has several key limitations. First, while our experimental design was motivated by a real-world context (bookreading) and child behavior (saying “again!” to demand a repeat read), the mechanics of the experiment—in which individual pages, rather than the entire book, repeat again and again—are of dubious ecological validity. Second, and relatedly, we introduced the DISTRACTOR to ‘catch’ children’s attention when it was no longer captured by the story and its illustration. However, the animation’s presence itself may have interfered with children’s attention to and processing of the story. Understanding the role of the DISTRACTOR in children’s learning is important for understanding what our method can tell us about how children attend to speech in the real world. Of course, children are virtually never offered a riveting .gif to watch, as an alternative to paying attention to a speech stream. However, children are arguably always faced with potential competitors for the attention that they could be devoting to language in their environments (e.g., when engaged with a younger sibling, rather than listening to their caregivers’ conversation, or even when just thinking really hard about dinosaurs, rather than hearing what someone has to say). Finally, we used child age as a proxy for children’s level of cognitive-linguistic development. This is defensible based on previous samples, which have shown child age to be highly correlated with vocabulary size, and moderately correlated with unfamiliar word generalization performance. However, future studies could more precisely test the hypotheses advanced here by employing language-specific measures of development, rather than relying on age as an index.

3.2 Conclusion

We designed a novel method to directly test the classic idea that children’s attention to a stimulus is driven by its support for their learning, and to extend this idea to an important new domain: spoken language. Our results reveal one of many ways in which children can be thought of as *active* language-learners (Bloom, 2000; Foushee et al., 2021, 2023; Saylor & Ganea, 2018; Zettersten & Saffran, 2021), effectively shaping their language input via the deployment of their own attention.

References

- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown* [R package version 0.1.0.9997]. <https://github.com/crsh/papaja>
- 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., 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>
- Bloom, L. (2000). The Intentionality Model of Word Learning: How to Learn a Word, Any Word. In R. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. Smith, A. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 19–50). NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195130324.003.002>
- Brown, R. (1973). A first language: The early stages. *Harvard University*.
- 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.
- Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of learning and memory*, 92(2), 225–234.
- 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*.
- Cubit, L. S., Canale, R., Handsman, R., Kidd, C., & Benvenuto, L. (2021). Visual attention preference for intermediate predictability in young children. *Child development*, 92(2), 691–703.
- Dunn, L. M., & Dunn, L. M. (1981). Peabody picture vocabulary test-revised.
- 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.) (2nd ed.). Baltimore, MD: Brookes.
- Forest, T. A., Siegelman, N., & Finn, A. S. (2022). Attention shifts to more complex structures with experience. *Psychological Science*, 33(12), 2059–2072.
- Foushee, R., Griffiths, T. L., & Srinivasan, M. (2016). Lexical Complexity of Child-Directed and Overheard Speech : Implications for Learning. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 1697–1702.
- Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, 206(July 2020), 104415. <https://doi.org/10.1016/j.cognition.2020.104415>
- Foushee, R., Srinivasan, M., & Xu, F. (2023). Active learning in language development. *Current Directions in Psychological Science*, 32(3).
- 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>
- Gureckis, T. M., & Markant, D. B. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, 7(5), 464–481. <https://doi.org/10.1177/1745691612454304>
- 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., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, 88(3), 449–460.
- 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-Gonzalez, H., & Brysbaert, M. (2012). *Age-of-acquisition ratings for 30, 000 {English} words*.
- 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.
- Poli, F., Serino, G., Mars, R., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science advances*, 6(39), eabb5053.

- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, 17(6), 880–891. <https://doi.org/10.1111/desc.12172>
- Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., et al. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*, 92(2), 135–138.
- 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>
- Saylor, M. M., & Ganea, P. A. (2018). *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms*. <https://doi.org/10.1007/978-3-319-77182-3>
- Shepherd, S. S., & Kidd, C. (2024). Visual engagement is not synonymous with learning in young children. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*.
- Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, 40(3), 672–686. <https://doi.org/10.1017/S0305000912000141>
- Shneidman, L. A., & Goldin-Meadow, S. (2012). Mayan and u.s. caregivers simplify speech to children. In A. Biller, E. Chung, & A. Kimball (Eds.), *Proceedings of the 36th annual boston university conference on language development* (pp. 536–544, Vol. 2). Somerville, MA: Cascadilla Press.
- 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. https://books.google.se/books?id=RxjjUefze%7B%5C_%7DoC
- Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
doi: 10.1177/0956797613488145.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.
- Zettersten, M., & Saffran, J. R. (2021). Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Developmental Science*, 24(3), 1–12. <https://doi.org/10.1111/desc.13064>