Developmental Psychology
2014, Vol. 50, No. 5, 1600–1613

0012-1649/14/$12.00 DOI: 10.1037/a0035591

Children and Adults Integrate Talker and Verb Information in Online Processing

Arielle Borovsky
Florida State University

Sarah C. Creel
University of California, San Diego

Children seem able to efficiently interpret a variety of linguistic cues during speech comprehension, yet have difficulty interpreting sources of nonlinguistic and paralinguistic information that accompany speech. The current study asked whether (paralinguistic) voice-activated role knowledge is rapidly interpreted in coordination with a linguistic cue (a sentential action) during speech comprehension in an eye-tracked sentence comprehension task with children (ages 3–10 years) and college-aged adults. Participants were initially familiarized with 2 talkers who identified their respective roles (e.g., PRINCESS and PIRATE) before hearing a previously introduced talker name an action and object ("I want to hold the sword," in the pirate's voice). As the sentence was spoken, eye movements were recorded to 4 objects that varied in relationship to the sentential talker and action (target: SWORD, talker-related: SHIP, action-related: WAND, and unrelated: CARRIAGE). The task was to select the named image. Even young child listeners rapidly combined inferences about talker identity with the action, allowing them to fixate on the target before it was mentioned, although there were developmental and vocabulary differences on this task. Results suggest that children, like adults, store real-world knowledge of a talker's role and actively use this information to interpret speech.

Keywords: sentence processing, language acquisition, individual differences, eye movements, speech comprehension

Supplemental materials: http://dx.doi.org/10.1037/a0035591.supp

The sentence "I'll take care of him" has entirely different connotations depending on who speaks the sentence (e.g., mother vs. mob boss), yet classic accounts of sentence processing rarely consider the influence of talker information. Adults fluently generate inferences based on this paralinguistic talker-specific information from everyday speech (e.g., Van Berkum, Van Den Brink, Tesink, Kos, & Hagoort, 2008), but little is known about whether and how children generate inferences about sentential meaning depending on the talker's role. In this study, we examined for the first time how adults and children integrate long-term knowledge about a talker's role into their interpretation of unfolding spoken sentences using a visual-world sentence comprehension task.

The ease with which we interpret speech in everyday conversation disguises the incredible complexity of this task. Listeners continuously interpret the sounds, meaning, and syntactic structure of the words in spoken sentences while considering higher level cues that extend beyond the immediate utterance, such as the referential, event, and discourse context. This includes a rich web of information about the talker, such as mood, age, familiarity, and preferences. Even more impressively, listeners interpret this information actively. That is, rather than passively waiting to receive new information from the unfolding speech stream, listeners develop predictions about words that are likely to be spoken next (Altmann & Kamide, 1999; Altmann & Mirkovic, 2009; Federmeier, 2007; Kamide, Altmann, & Haywood, 2003). This predictive ability is even more remarkable considering that some information available in the speech stream is not entirely relevant for identifying words or syntax (e.g., voice fundamental frequency [f0]). Therefore, from an early age, listeners must not only interpret this rapid stream of spoken information *predictively*, but must also attend *selectively* to speech cues that are relevant for interpretation while disregarding irrelevant information.

How do listeners learn to identify which cues are relevant for speech interpretation? Until now, many studies in the adult literature have examined which sources of linguistic information in the speech stream are used to interpret sentences (see Huettig, Rommers, & Meyer, 2011, for a review). In many cases, these studies focus on the timing with which adults interpret temporary ambiguities in sentence structure or identify objects in a visual scene that have not yet been mentioned. This enterprise has been mirrored in the developmental literature, with findings that children are sensitive to a plethora of linguistic cues in real-time language

This article was published Online First March 10, 2014.

Arielle Borovsky, Department of Psychology, Florida State University; Sarah C. Creel, University of California, San Diego.

This research was supported by grants from the National Institutes of Health (F32-DC10106 to Arielle Borovsky) and from the National Science Foundation (CAREER Award BCS-1057080 to Sarah C. Creel). We are grateful to the families and schools that participated in this study. We wish to thank Heather Pelton for her assistance with collecting the norming data at preschools and Kim Sweeney and the research assistants at the Center for Research in Language for their help with the data collection.

Correspondence concerning this article should be addressed to Arielle Borovsky, Department of Psychology, Florida State University, 1107 West Call Street, Tallahassee, FL 32306-4301. E-mail: borovsky@psy.fsu.edu

interpretation, such as verbs (Fernald, Zangl, Portillo, & Marchman, 2008; Mani & Huettig, 2012; Nation, Marshall, & Altmann, 2003), adjectives (Fernald, Thorpe, & Marchman, 2010; Huang & Snedeker, 2009), gendered articles and pronouns (Arnold, Brown-Schmidt, & Trueswell, 2007; Lew-Williams & Fernald, 2007), and structural cues (Snedeker & Trueswell, 2004).

In the current study, we extended this research in a novel direction. We asked how adult and child listeners integrate linguistic information with cues that are not typically considered in linguistic models of sentence interpretation. Specifically, we asked how listeners use a speaker's voice to integrate pre-existing knowledge about that speaker in coordination with a verbal cue. Unlike in linguistic cue integration, children seem to achieve adult-like sensitivity to nonlinguistic and paralinguistic cues over a protracted developmental timescale. Developmental differences have been noted in the processing of referential context (Kidd & Bavin, 2005; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999), emotional prosody (Morton & Trehub, 2001; Quam & Swingley, 2012), and subtle acoustic cues to talker identity (Creel & Jimenez, 2012; Mann, Diamond, & Carey, 1979). These findings motivate a need to examine how children and adults acquire the ability to integrate paralinguistic information with linguistic cues in the speech stream during real-time speech comprehension.

Why Talker Identity?

Nonlinguistic and paralinguistic cues are not typically mentioned in traditional models of speech comprehension. Yet, it is clear that attention to nonlinguistic detail that is traditionally considered irrelevant or even detrimental to word recognition can potentially facilitate speech interpretation by providing semantic cues to the utterance (Creel & Bregman, 2011). For example, in English, speaker variation in vocal pitch does not distinguish different word meanings but may help the listener recognize the talker. In turn, knowing the talker's preferences may aid in semantic interpretation of the speech signal.

Adult listeners make numerous inferences about the talker's characteristics using so-called *indexical cues* in the speech signal (Abercrombie, 1967; Ladefoged & Broadbent, 1957; Peirce, 1903/1998), including judgments about sex and age (Peterson & Barney, 1952), height and weight (Krauss, Freyberg, & Morsella, 2002), social class (Labov, 1972), and region of origin (Clopper & Pisoni, 2004). Less is known about whether and how children make similar inferences about a talker's characteristics using such indexical cues. Some evidence suggests that young children can make some social decisions based on spoken-language cues to identity. For example, 5-year-olds prefer to be friends with someone who speaks the same language or accent as they do (Kinzler, Dupoux, & Spelke, 2007; Kinzler, Shutts, DeJesus, & Spelke, 2009), and preschoolers also ascribe less familiar objects to foreign-language voices (Hirschfeld & Gelman, 1997). However, more fine-grained aspects of voice recognition improve into adolescence, such as distinguishing between two unfamiliar voices of the same gender (Bartholomeus, 1973; Mann et al., 1979; Spence, Rollins, & Jerger, 2002). Developmental differences in perceptual discrimination of talker identity might lead to downstream consequences for cognitive and linguistic processing of the speech stream.

Young children can use broad acoustic distinctions between talkers to facilitate sentence processing (Creel, 2012). Creel presented preschool-age children and adults with information about each of two talkers' color preferences. In various experiments, participants were told that a female character, Anna, preferred one color (e.g., pink or black), while a male character, Billy, preferred another color (e.g., blue or white). Creel found that this simple, explicitly instructed talker information is rapidly activated in both children and adults. When a talker's color preference was repeatedly mentioned, listeners subsequently looked to items of that color when hearing that talker's voice again. Further, children even looked to items of the character's preferred color *when mentioned by the other talker* (e.g., Billy saying "Anna wants to see the square" elicited looks to pink pictures), suggesting that when children used talker information, they were using it to infer which character wanted to see the shape. These findings indicate that, at least in a simple task where talkers' preferences are explicitly mentioned and repeatedly highlighted, children use voice cues to constrain their online interpretations of sentences much as adults do.

Integrating Talker Identity and Event Knowledge

It is not clear if children immediately activate information that has not been recently mentioned about a speaker during online sentence processing, nor is it known whether this information could be readily combined with downstream linguistic cues, such as a sentential action. Adult listeners seem to activate world knowledge about sentential agents immediately during sentence interpretation. For example, listeners use role-typical information (in combination with a verb) to resolve thematic reference during sentence processing (Kamide, Altmann, et al., 2003). Adult listeners can also clearly integrate this real-world knowledge with downstream verbal and morphosyntactic cues such as case and tense marking when developing predictions about the likely outcome of an event (e.g., by looking to an unmentioned empty or full glass when hearing "a man has drunk" or "will drink," respectively; Altmann & Kamide, 2007; Kamide, Scheepers, & Altmann, 2003). Like adults, children can also integrate explicitly named agents with other sentence constituents. Borovsky, Elman, and Fernald (2012) presented adults and children (ages 3–10 years) with simple spoken sentences such as "The pirate chases the ship." Even the youngest children combined a sentential agent (like *pirate*) with a sentential action (like *chases*) to develop online predictions about subsequent items in the sentence (e.g., the object of the sentence, like SHIP). These findings indicate that children as young as 3 years can rapidly integrate linguistically specified information about agents and actions to update expectations about upcoming items dynamically.

However, it is an open question as to whether children integrate paralinguistic cues, such as that of a talker's inferred *identity*, with linguistic cues during speech processing (i.e., hearing a piratical voice say "I want to chase the ship"). Some evidence suggests that adults integrate talker cues with sentence structure to generate online inferences about upcoming spoken language. Specifically, Van Berkum et al. (2008) found that adults showed a larger semantic evoked potential (N400) when the combination of a talker's voice and sentence content violated social norms (e.g., a child talker saying "I want to drink wine") than when social norms were not violated. This research indicates that semantically relevant voice information influenced adults' processing at least by the word of interest.

The Current Study

While prior work demonstrates that both adult and child listeners rapidly activate information about talkers' preferences from voice information, a number of unanswered questions remain. First, do children, like adults, rapidly activate world knowledge about talkers from voice information alone, or are children's abilities limited to simple, dichotomous distinctions (e.g., talkers' preferences for different colors explicitly taught in a lab setting; Creel, 2012)? Second, Van Berkum et al.'s (2008) study with adults does not make clear *when* talker identity cues are integrated with the sentence. Did the listeners simply recognize an incongruity between the target word and the talker's identity (i.e., that children rarely semantically co-occur with wine), or did listeners anticipate likely events that a child would participate in (i.e., the entire event was incongruous), or both? These questions formed the focus of the current study.

A secondary question concerns the relationship between language skills and cue integration generally. Individual differences in linguistic abilities, such as vocabulary, have been repeatedly shown to affect the speed with which very young children interpret words and sentences (Borovsky et al., 2012; Fernald, Perfors, & Marchman, 2006; Mani & Huettig, 2012; Marchman & Fernald, 2008). Intriguingly, recent evidence has suggested that individual differences in linguistic processing may extend into adulthood. For example, measures of adult literacy and vocabulary skills associate with speed and proficiency of interpretation of both words (Huettig, Singh, & Mishra, 2011) and sentences (Borovsky et al., 2012; Mishra, Singh, Pandey, & Huettig, 2012). Yet no work has yet examined how individual differences in vocabulary skill may relate to the integration of talker cues in online speech processing in children or adults. One possibility is that the use of talker information may pattern with other perceptual talker-discrimination abilities and might simply improve with age, (Creel & Jimenez, 2012; Mann et al., 1979). Alternatively, if talker information serves as one of many cues to sentence meaning, then general language processing abilities, as measured by age-relative vocabulary skill, might pattern with the ability to rapidly integrate talker cues with sentential information. Therefore, we explored how vocabulary ability interacts with integration of talker and event representations in sentences in both children and adults.

To test our hypotheses, we used the "visual-world" eye tracking paradigm (VWP; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), which measures moment-to-moment changes in participants' gaze to objects in a visual scene in response to spoken language (see Huettig, Rommers, et al., 2011 for a review). Changes in participants' visual fixations can reflect real-time anticipation of upcoming information (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). It is important to note that the VWP does not require a speeded manual response from participants, and performance on these and similar eye-tracking tasks have been sensitive to individual differences in adult and child vocabulary (Borovsky et al., 2012; Fernald et al., 2006; Mani & Huettig, 2012; Marchman & Fernald, 2008), so it is well suited for measuring rapid processing of ongoing speech in young children and adults.

The current task (see Figure 1) is a modification of Borovsky et al. (2012). Borovsky and colleagues presented children (ages 3–10 years) and adults with sentences ("The pirate will chase the ship") consisting of a subject noun (PIRATE), an action (CHASE), and an object (SHIP). Each sentence was heard while listeners viewed pictures related to the subject (SHIP ["chase-able" by a pirate], TREASURE [something hidden by a pirate]) and to the action (SHIP, CAT [chase-able by a dog, the subject noun on other trials]), as well as an unrelated picture (BONE [something hidden by a dog]). Even the youngest children fixated the ship (the only pirate-related, chase-able picture) before they heard the word *ship*. In the current study, the subject noun (PIRATE) was replaced by *identity implied by the voice*. Thus, participants might hear the "pirate" voice say "I want to hold the sword," while viewing four items related to the talker (SWORD, SHIP) and the action (SWORD, WAND), along with an unrelated item (CARRIAGE). If listeners can interpret talker cues immediately, then they should look toward talker-related items (SWORD, SHIP) as soon as they can identify the talker (during "I want to"). If they can use action cues immediately (as in Borovsky et al., 2012), they should look more to SWORD and WAND shortly after hearing the action word. If they can combine these cues, then the greatest proportion of looks should be toward the target item (SWORD)—the only one that is *both* talker-related and action-related. However, if listeners fail to consider talker identity, then looks to the target item should not differentiate from looks to the action-related competitor until after the target word is spoken.

Method

Participants

Forty-nine native English-speaking undergraduate students (mean age: 21.3 years, 38 women) took part in the experiment for course credit. All students were enrolled at a public research university located in southern California in the San Diego metropolitan region. Although we specifically recruited native-English-speaking participants who had reported learning only a single language before the age of 6 years, our sample was diverse; 49% of the population reported belonging to an ethnic minority. Participant socioeconomic status was not requested, although other work on the undergraduate population has indicated that the students comprise a range of childhood socioeconomic statuses (SES), though the majority of students were raised in middle to upper SES households. Participants reported normal hearing and vision and no history of mental illness or treatment for speech, language, or cognitive issues. Six more participated but were excluded from analysis for the following reasons: two reported they were familiar with the aims of the study, two participants had excessive data loss (more than 50% of gaze samples in all trials not captured), one had glaucoma, and one had received speech therapy.

Forty-nine monolingual English-learning children (21 girls) between the ages of 3 years 0 months and 10 years 0 months (mean age: 7 years 1 month) were recruited from the surrounding metropolitan region (San Diego, CA). Children had either previously participated in child language research or their par-

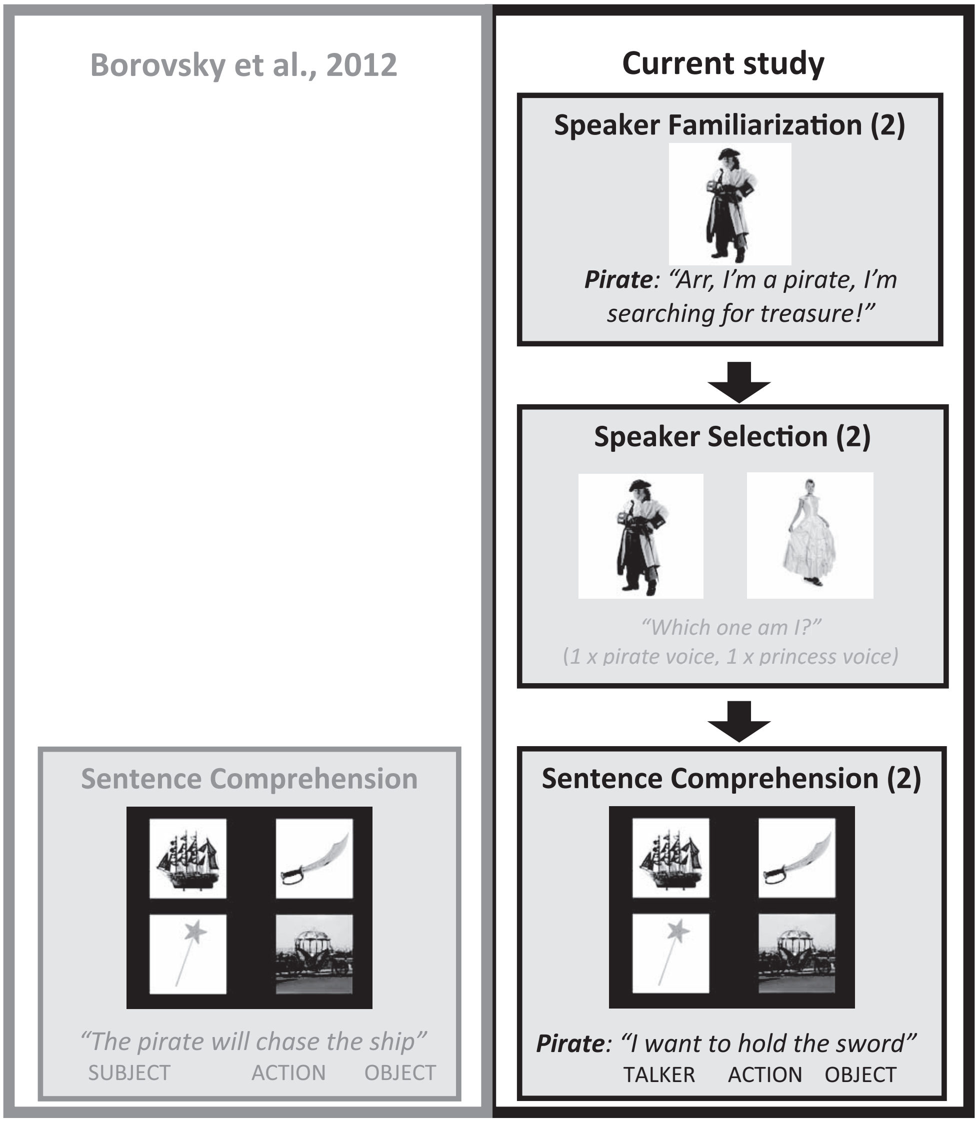


Figure 1. Schematic depiction of experimental procedure, compared with procedure of Borovsky et al. (2012). In the current study, the depicted trial sequence repeated six times, for a total of 12 trials in each subtask. Note: Similar, but not identical, images were used in actual study; they were not reprinted here due to copyright restrictions. Pirate, sword, carriage, ship, and wand images courtesy of http://www.istockphoto.com; princess image courtesy of http://www.canstockphoto.com.

ents had answered invitational flyers and ads posted in the community. The resulting sample consisted of 27% children of ethnic minority and 6.1% from single-parent households; all mothers had completed at least high school and reported an average 16.5 years of education. Each child received a toy in return for participation, and their families received $10. Parental report indicated that each child participant had normal hearing and vision, was hearing primarily English at home, was typically developing, did not experience significant birth complications, had no recent or chronic ear infections, and had received no diagnosis or treatment for other language, speech, motor, or cognitive issues. Twelve additional children also participated but failed to meet one or more of these criteria. One additional child was excluded who completed fewer than 50% of trials in the experiment.

Stimuli

The experiment consisted of two phases within each block, *talker familiarization* and *sentence comprehension*, with stimuli designed for each phase. In the talker familiarization phase, role images (e.g., PIRATE; see Figure 1) were paired with voices that introduced each role to the participant, with minimal accompanying semantic information (e.g., "Arrr! I'm a pirate! I'm searching for treasure!"). Roles and accompanying 400 $\times$ 400 pixel images were selected to be highly familiar and identifiable to preschool children (as confirmed by norming, described in the online supplemental materials) and to have distinct, stereotypical voices (e.g., PRINCESS vs. PIRATE).

The six sentence quartets (Appendix) were developed such that each role member in a pair (e.g., PRINCESS and PIRATE) would elicit different object expectations (e.g., WAND or SWORD) when paired with the same verb (HOLDS). It should be noted that none of the objects mentioned during talker familiarization occurred as targets in the sentence comprehension trials. For instance, *wand*, *sword*, *ship*, and *carriage* were not mentioned during talker familiarization. This means that to predict targets, listeners had to activate long-term semantic associations with, for instance, pirates and princesses. Sentence quartets were composed by crossing both roles with two actions to create four sentences, with two spoken by each talker. We chose 400 × 400 pixel images that corresponded to the object of each sentence in the quartet (see Borovsky et al., 2012). Therefore, each visual scene consisted of four images that varied in their relationship to the sentence and the two possible talkers. Each image served as (a) a target (corresponding to the sentential object) or one of three types of competitors: (b) talker-related, (c) action-related, or (d) unrelated. Because four sentences were associated with each image quartet, each image served in each condition an equal number of times, yielding a balanced within-subject design.

Twelve native English speakers (six female, six male) served as voices for the 12 roles in the study. Voice pairs varied by gender because prior research suggests that preschool age children have difficulty distinguishing between same-gender talkers due to their greater acoustic similarity (Creel & Jimenez, 2012).

Auditory stimuli were edited using Praat software (Boersma & Weenink, 2012) and were adjusted to a standard mean intensity (70 dB). Sounds for the sentence comprehension phase were also normalized to standard duration (see the online supplemental materials for details). Aligning critical-word onsets allowed us to average across all sentences without time distortions, so that the time course of looks based on each critical word type—particularly, talker cues (first available during *I want to*), action (*hold*), and theme (*sword*)—was directly evident.

Procedure and Equipment

Stimuli were presented on a PC computer running Experiment Builder software (2011). Participants were first seated in a comfortable stationary chair. We individually adjusted the position of a 17-in. display on a remote arm to fit the participant's height and to ensure that the display and eye-tracking camera were placed 580–620 mm from the participant's face. Eye movements were automatically tracked by an Eyelink 2000 remote eye tracker (2004) at 500 Hz. A target sticker affixed to the participant's forehead allowed the eye tracker to maintain a stable measurement of the participant's gaze despite head movements. Next, the eye tracker was calibrated with a 5-point routine using a standard black-and-white 20-pixel-diameter bull's-eye image. This dot also served as a drift-correction dot prior to recording eye movements for each trial in the sentence comprehension task.

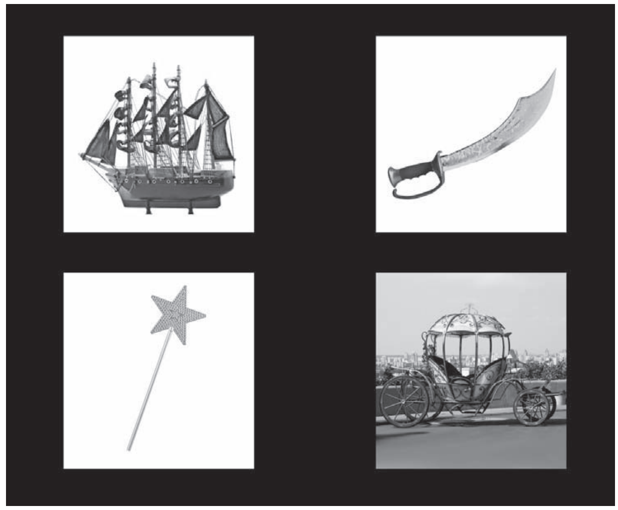
Participants were then instructed that they would be seeing pictures and listening to different characters. They were asked to listen carefully to each character and to select the picture that went best with the sentence by pointing to the image (children) or clicking on the image with the mouse (adults). They were then given an example of the sentence comprehension task. All participants immediately selected the target image corresponding to the sentence final object, and no further practice was required. Participants were then instructed to focus on the bull's-eye image whenever it appeared in a central location when they were ready to proceed.

The experiment consisted of six cycles of three interleaved trial types (Figure 1): two *talker familiarization* trials, where two talkers were introduced; two *talker selection trials*, to verify that children mapped voices to roles; and finally, two *sentence comprehension* trials where the participants selected images that corresponded to two sentences. During talker familiarization trials, participants first viewed a single role image on the screen (e.g., PIRATE) for 2,000 ms, before the talker introduced their role. This procedure was then repeated for a second role. Next, both role images reappeared side by side. Each talker asked, "Which one am I?" Participants selected the appropriate role by either pointing to the appropriate role image (children) or selecting it with a mouse (adults).

Then, the participant completed two sentence comprehension trials (Figure 2). Participants were first asked to look to a central fixation point. Next, four images appeared on the screen for 2,000 ms before sentence onset. In each sentence, they heard one of the two prefamiliarized characters say, "I want to [ACTION] the [OBJECT]" (e.g., Pirate: "I want to hold the sword"). The object corresponded to one of the four images in the array. The talker's role was not mentioned or illustrated during this task, meaning that the participants had to recognize the character from his or her voice alone. After the sentence ended, a mouse icon appeared at the center of the screen, and the images remained until the participant indicated their image selection (by pointing or clicking). Fixations were recorded in each trial from the onset of the images until picture selection occurred. A second sentence followed. After the two-sentence comprehension trials, another cycle of familiarization and comprehension trials began. Recalibration of the eye tracker was performed if needed, although this was rarely necessary. Participants were given a break halfway through the study (after three blocks). The entire task took 5–10 min.

Design. The visual and auditory stimuli for the sentence comprehension task were counterbalanced across and within experimental versions. In each version of the study, two out of the four possible comprehension sentences (see Appendix) were pseudorandomly selected to appear, such that across participants, all possible sentence combinations appeared with equal frequency. Additionally, every image appeared with equal frequency in each quadrant, and all combinations of the target and competitor image positions were balanced across all versions. Within any version of the study, the target and competitor pictures appeared in all quadrants with equal frequency. The counterbalancing scheme aimed to minimize biases to view or select particular objects in a particular screen location.

Data processing. Offline, visual fixation data were binned into 10-ms intervals, over which subsequent analyses were performed. We removed trials where the participants failed to attend to the computer screen for the majority of the sentence (fewer than 50% of the samples were measured in fixation to any region of the screen). This resulted in the exclusion of two



	Target	Talker-Related	Action-Related	Unrelated
Princess voice:				
I want to hold the wand.	WAND	CARRIAGE	SWORD	SHIP
I want to ride in the carriage.	CARRIAGE	WAND	SHIP	SWORD
Pirate voice:				
I want to hold the sword.	SWORD	SHIP	WAND	CARRIAGE
I want to ride in the ship.	SHIP	SWORD	CARRIAGE	WAND

Figure 2. An example of a display in the experiment and four possible spoken sentences paired with this display. A given participant only heard two of the four sentences possible. Note: Similar, but not identical, images were used in actual study but were not reprinted here due to copyright restrictions. Sword, carriage, ship, and wand images courtesy of http://www.istockphoto.com.

trials for adult participants (0.5%) and 14 trials for the children (2.5%).

Vocabulary assessment. In prior work, vocabulary level predicted anticipatory sentence processing in children and adults (Borovsky et al., 2012). Therefore, after the completion of the eye-tracking task, participants were administered an offline measurement of vocabulary level: the Peabody Picture Vocabulary Test, Version 4 (PPVT-4; Dunn & Dunn, 2007). The PPVT-4 is a standardized test of receptive vocabulary skill in participants between the ages of 2 and 90 years. For our analyses, we age-adjusted raw PPVT scores prior to analysis by calculating residualized vocabulary when regressed against participant age. This method yields a standardized measure of each participant's vocabulary ability that is controlled for linear effects of age, unlike a measure of raw vocabulary size. Borovsky et al. (2012) previously found this residualized vocabulary measure to be a predictor of sentential cue integration in adults and children, and this procedure has been adopted in a variety of language acquisition and processing research, as it provides a continuous, normalized measure of vocabulary ability with respect to the experimental sample (e.g., McAuliffe, Gibson, Kerr, Anderson, & LaShell, 2013; Shafto, Geren, & Snedeker, 2010).

Results

We measured both behavioral accuracy of target picture selection and visual fixations to target and competitor pictures throughout the task. Accuracy was high throughout, verifying that participants understood the task. Visual fixations were analyzed more extensively, as they provide information about the time course over which talker and sentence cues were integrated in comprehension.

Accuracy Analyses

We first asked whether adults and children associated the talker with the appropriate role by measuring accuracy in the talker selection task. We then verified that sentence comprehension was accurate. As expected, accuracy on role selection was extremely high: adults had perfect accuracy (100%), and children had near-perfect accuracy (98.8%; seven total errors). All errors were made on the first trial of the block, and none of the children selected the incorrect role twice in a block, suggesting they had inferred the appropriate talker–role association by the end of the block. Therefore, no trials were removed on the basis of behavioral performance in the talker selection task. We tested whether participants understood the sentence comprehension task by measuring accuracy of target picture selection. Adults selected the target picture 100% of the time, and children did so 98.8% of the time (seven total errors). These error trials were removed prior to eye-tracking analyses, as were trials where the participants failed to attend to the computer screen for the majority of the sentence. We excluded two trials for adult participants (0.5%) and 14 trials for the children (2.5%) using this criterion.

Visual Fixation Analyses

Our main analyses focused on the moment-by-moment visual fixations toward the four experimental interest areas. We first describe the general time course of eye movements in children and adults. Next, we quantify how age and vocabulary ability influenced the fixations to the target item versus the other items.

Figure 3 illustrates adults' and children's real-time interpretation of the test sentences. Several fixation patterns are evident in response to the unfolding events of the sentence. As talker identity cues became available—that is, as the speaker uttered "I want to . . . " (from sentence onset to 1,068-ms post-onset)— listeners began to fixate the two talker-related items (e.g., SWORD, SHIP). This pattern suggests that our listeners succeeded in using the talker's voice to generate appropriate (predictive) inferences about the talker's preferences. Next, as the verb and article were uttered

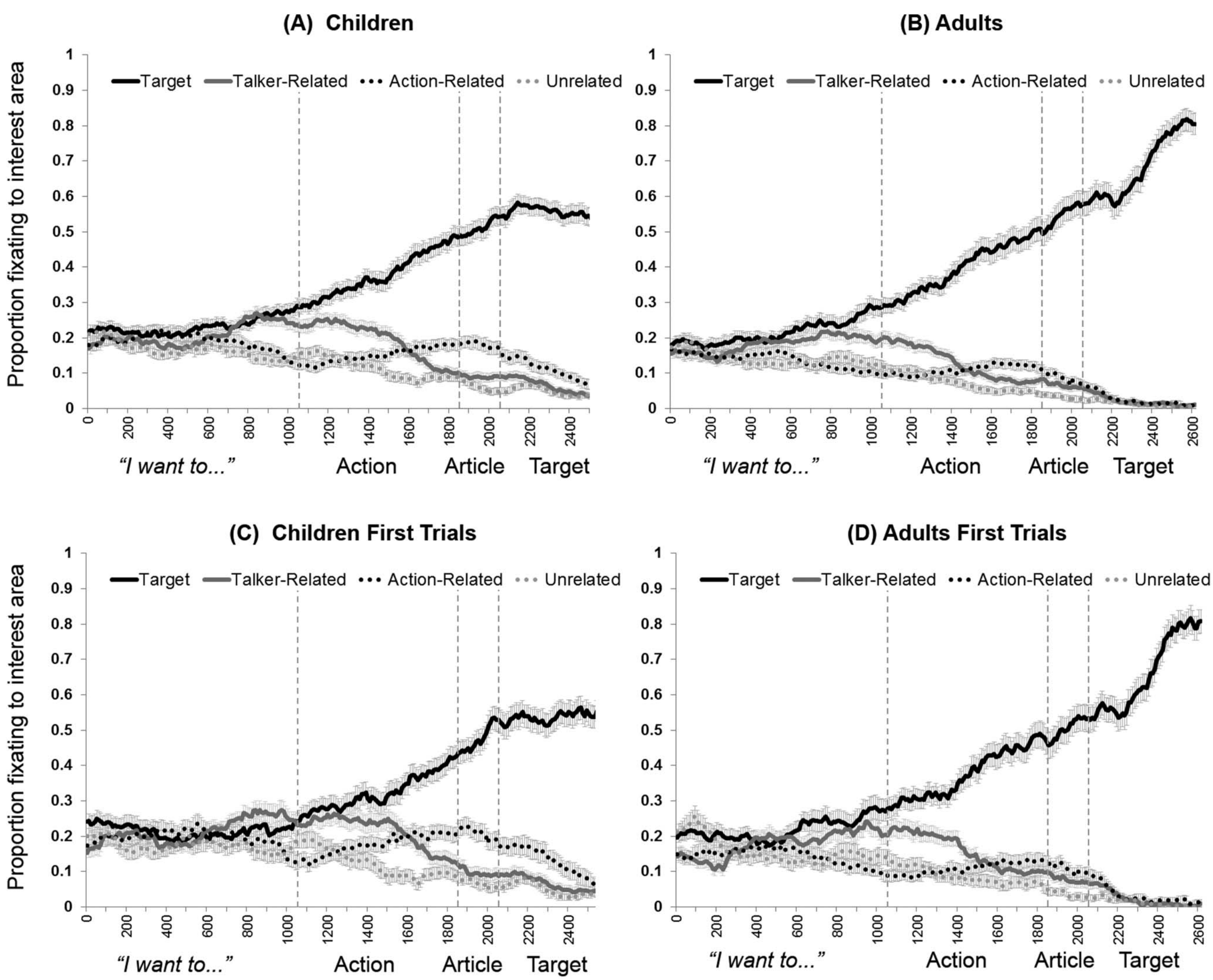


Figure 3. Time course plots of mean proportion of fixations and standard errors in 10-ms intervals toward the target and competitor items from sentence onset to offset in (A) children (ages 3–10 years) and (B) adults and for the first trials of each block in (C) children and (D) adults. Vertical dashed lines indicate onset times of the words of interest.

(e.g., "hold the . . . "; from 1,068- to 1,928-ms postsentence onset), listeners looked more to both of the action-related items (SWORD, WAND) than to the unrelated distractor. Additionally, soon after the verb was mentioned, listeners combined the talker-related cues with the verb to fixate primarily toward the target item (e.g., SWORD). This pattern indicated that listeners updated their fixations to correspond to locally relevant information (the action word) as it was spoken. These fixation patterns correspond closely to prior studies of simple sentence processing in children and adults (Borovsky et al., 2012; Fernald et al., 2010; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). Thus, overall, both children and adults appear to integrate talker-identity information with their event knowledge to develop dynamic real-time predictions about their preferences.

The reader may notice that target looks in Figures 3A and 3B began to increase over other looks prior to action-word onset, which should not be possible based on information in the sentence. This apparent "precognition" effect might result from the trial design, in which listeners heard two successive sentence comprehension trials with two different targets. If listeners select the sword as the target on Trial 1, they may disfavor the sword as a possibility on Trial 2. This would artificially heighten looks to all non-sword pictures, including the target. A greater proportion of target looks on the second trial was confirmed in analyses of target fixations (discussed later). For first trials alone (Figures 3C and 3D), "precognition" was absent—fixations to the target did not exceed looks to the talker-related picture until *after* action-word onset. Nonetheless, the pattern of responses to talker cues and to action-word cues resembled the full set of trials. We included trial (first, second) as a factor in all following analyses, for caution's sake. However, trial never interacted with other factors.

The main goal of this research was to investigate whether listeners could actively generate predictive inferences about the likely target object by simultaneously integrating the talker's role (from the talker's voice) in combination with the spoken sentential action. If predictive processing occurred, we would expect the magnitude of fixations toward the target to have exceeded looks to all other competitor images before the sentential object is spoken, but after both cues are available to the listener. Therefore, we quantified predictive looking as the mean proportion of time spent fixating to the target object versus each of the other objects during the anticipatory time window, spanning action onset to target onset. In this time window, we computed *target advantage scores* as the difference between the proportion of anticipatory fixations to the target minus each of the other pictures types. This yielded three difference scores: (a) target–talker-related, (b) target–action-related, and (c) talker–unrelated. A target advantage of 0 reflects equal anticipatory looking proportions to target and a competitor, whereas a positive target advantage reflects more anticipatory looks to the target. If listeners are using talker cues, they should look more to the target than to the action competitor and the unrelated competitor during the anticipatory window. Further, if they are using action cues, they should look more to the target than to the talker-related and unrelated competitors. If they are not using either cue, then there should be no anticipatory fixation differences until *after* the onset of the target word—that is, all target advantages should be 0.

We were additionally interested in the influence of developmental differences and vocabulary level on sentence processing. Therefore, we included age-normalized vocabulary ability (described previously in Methods) and age as factors in our analyses. Figure 4 illustrates the time course of processing for older and younger children (Figures 4A and 4B) and for individuals with higher and lower vocabulary scores in adult (Figures 4C and 4D) and child (Figures 4E and 4F) groups. We explored developmental effects in two separate sets of analyses: (a) The first set of analyses included an age factor with adults and children as separate levels, and (b) the second set of analyses was restricted to children only, where the child age factor was split into older and younger child levels that were defined according to median split of participant age (older age range = from 7 years 3 months to 10 years 10 months; younger age range = from 3 years 0 moths to 7 years 1 month). This approach allowed us to initially explore developmental effects on our data as macro-level difference between adults and children across a wide age span and to subsequently measure finer-scale developmental differences between younger and older children. Higher and lower vocabulary ability groups were determined by median splits of age-normalized child and adult group vocabulary scores.

In the first analysis that included adults and children together, we carried out three mixed-model analyses of variance (ANOVAs), one on each target advantage score, with between-subjects factors of age (adult vs. child) and vocabulary (higher vs. lower) and the within-subject factor of trial order (first vs. second). An effect of age indicated that adults showed larger target advantage than children in all analyses: target–talker-related: $F(1, 93) = 6.92$, $p = .01$, $\eta_p^2 = .069$; target–action-related: $F(1, 93) = 5.68$, $p = .019$, $\eta_p^2 = .058$; target–unrelated: $F(1, 93) = 4.59$, $p = .035$, $\eta_p^2 = .047$. Vocabulary also significantly affected anticipatory fixation proportions: listeners with higher vocabulary skill showed larger differences than those with lower vocabulary skill: target–talker-related: $F(1, 93) = 7.38$, $p = .008$, $\eta_p^2 = .073$; target–action-related: $F(1, 93) = 5.17$, $p = .025$, $\eta_p^2 = .053$; target–unrelated: $F(1, 93) = 6.82$, $p = .011$, $\eta_p^2 = .08$. Finally, trial affected anticipatory target looks: participants fixated more to the target on the second (vs. first) trials in each block: target–talker-related: $F(1, 93) = 27.53$, $p < .0001$, $\eta_p^2 = .23$; target–action-related: $F(1, 93) = 21.83$, $p < .0001$, $\eta_p^2 = .19$; target–unrelated: $F(1, 93) = 21.43$, $p < .0001$, $\eta_p^2 = .19$. This confirms that listeners may have eliminated the first trial's target as a potential target on the second trial. However, trial order did not interact with age or vocabulary. Follow-up *t* tests found that target advantage scores were significantly above zero for both first and second trials for all competitor picture types, for both adults and children (Table 1).

We also explored whether there were developmental changes across age at a finer level. We conducted another set of three mixed-model ANOVAs with child data only, with vocabulary and child age (older children, younger children) as between-subjects factors and trial as a within-subject factor. Here, the magnitude of target advantage scores for children with higher vocabulary skills was marginally larger than those of children with lower vocabulary skills for the target–talker-related, $F(1, 45) = 3.93$ $p = .054$, $\eta_p^2 = .080$, and target–unrelated scores, $F(1, 45) = 3.35$ $p = .074$, $\eta_p^2 = .069$. Once again, the trial factor was significant: anticipatory target fixation proportions relative to all competitor pictures were

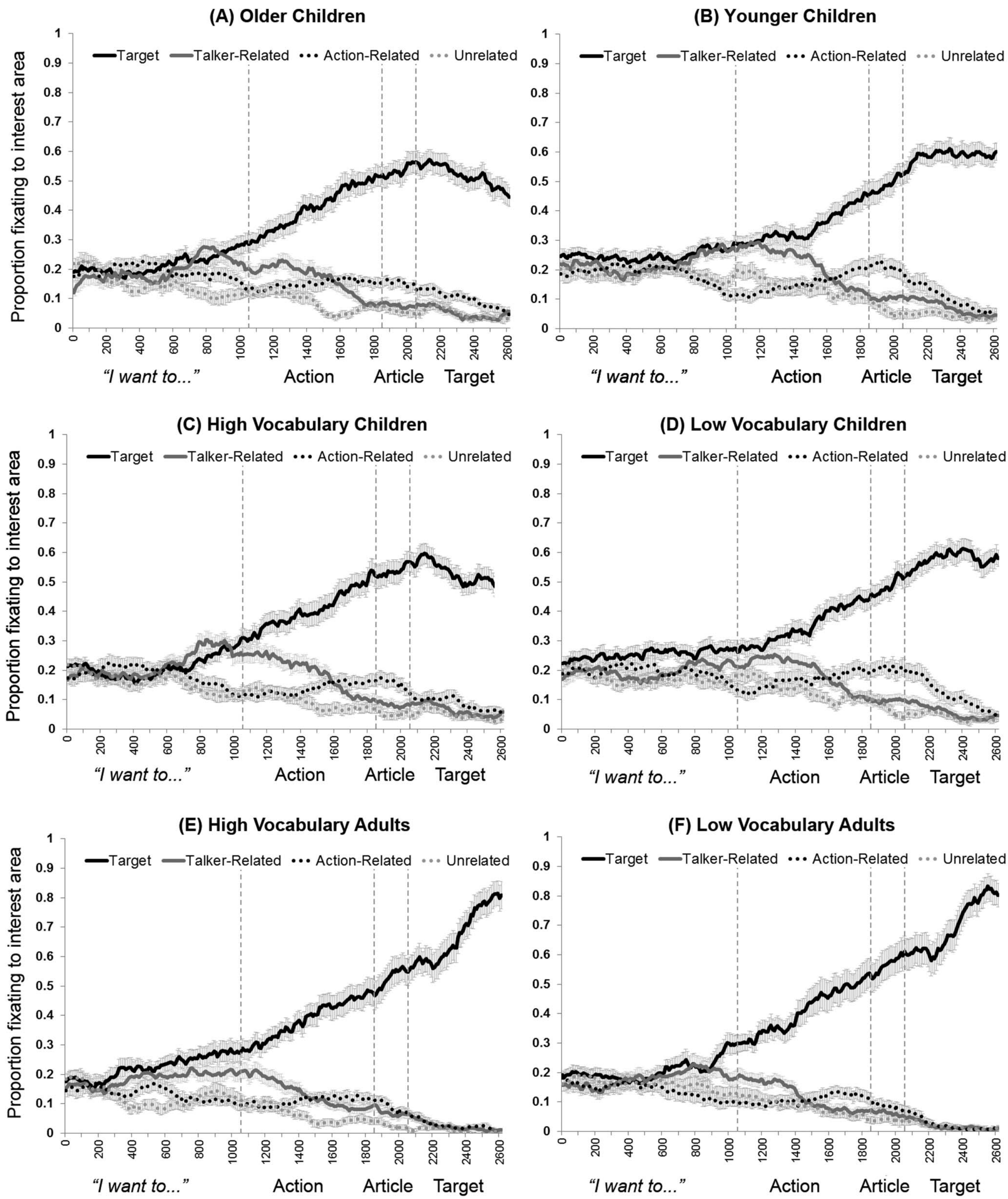


Figure 4. Time course plots of mean proportion of fixations and standard errors in 10-ms intervals toward the target and competitor items from sentence onset to offset in (A) older and (B) younger children; (C) high- and (D) low-vocabulary children; and (E) high- and (F) low-vocabulary adults.

Table 1
Target Advantage Scores for Fixation Proportions to Target Versus. Each of the Other Pictures During the Anticipatory Time Window

Variable	Target–talker-related	Target–action-related	Target–unrelated
All adults	.36	.38	.42
First trials	.29	.34	.38
Second trials	.43	.42	.47
All children	.26	.28	.33
First trials	.18	.20	.27
Second trials	.34	.36	.40

Note. Cohen's *d* statistics on these comparisons all indicated very strong effect sizes and ranged from 1.35 (children's first trials, target–talker-related) to 4.3 (all adults, target–unrelated). All $ps < .0001$.

larger for second versus first trials—target–talker-related: $F(1, 45) = 10.32$ $p = .0024$, $\eta_p^2 = .19$; target–action-related: $F(1, 45) = 16.25$, $p = .0002$, $\eta_p^2 = .27$; target–unrelated: $F(1, 45) = 11.14$ $p = .002$, $\eta_p^2 = .20$. There was no effect of child age[1]—target–talker-related: $F(1, 45) = 1.92$ $p = .17$, $\eta_p^2 = .041$; target–action-related: $F(1, 45) = 0.32$, $p = .57$, $\eta_p^2 = .007$; target–unrelated: $F(1, 45) = 1.45$ $p = .24$, $\eta_p^2 = .031$. No other effects or interactions, including those involving trial, approached significance.

Discussion

While existing evidence has suggested that adults can successfully integrate voice and role cues during spoken language interpretation (Van Berkum et al., 2008), it was not known whether children or adults could fluidly integrate voice cues with sentence structure to make inferences in real-time. We found that both child and adult listeners do so adeptly: 3- to 10-year-old children successfully activated talker-role-associated knowledge that was not mentioned in the experiment (for instance, that pirates like to hold swords), and they used this information to guide their interpretation of the talker's utterances. On hearing a talker's voice, children and adults looked (predictively) toward items that were associated with the talker's previously mentioned role. Then, as listeners encountered additional information in the sentence (the action), they successfully integrated the talker's role with the action, as indicated by a dramatic increase in the proportion of anticipatory fixations toward the likely target object before its label was spoken. This ability improved across a wide age range. Adults more strongly generated real-time inferences than children (ages 3–10 years), although older children (from 7 years 3 months to 10 years 10 months) did not differ from younger children (from 3 years 0 months to 7 years 1 month) on this task. Performance on this task was additionally linked with vocabulary ability such that participants who had larger vocabularies for their age generated predictive fixations more robustly than those with lower vocabulary skill.

These findings suggest that child listeners (ages 3–10 years) can and do integrate talker identity immediately into language interpretation and are consistent with an event-integration account of Van Berkum et al.'s (2008) event-related potentials study with adults. It also indicates that skill in talker–role cue integration improves across development and that the child's vocabulary skill is important in rapidly activating talker–role knowledge during sentence interpretation. In the remainder of the article, we discuss implications of our findings for language development and sentence processing and outline potential research directions of interest.

The Role of Vocabulary Knowledge

Our results are consistent with several accounts regarding the role of vocabulary development in language processing. One potential explanation for our observed vocabulary effects is that individuals who have larger vocabularies for their age may simply be more adept at language processing than those with smaller vocabularies. A number of eye-tracking studies have found associations between speed of word and sentence interpretation and vocabulary ability in age groups ranging from infancy to adulthood (Borovsky et al., 2012; Fernald et al., 2006; Mani & Huettig, 2012). It is debated whether this correlation between vocabulary and processing skill is driven by early differences in general processing ability that facilitate later vocabulary learning (Fernald & Marchman, 2012) or if differences in vocabulary knowledge (potentially mediated by differences in parental speech or early learning environments) may drive speed of lexical recognition (Fernald, Marchman, & Weisleder, 2013; Hurtado, Marchman, & Fernald, 2008). An alternative explanation for the vocabulary effects in our study may be that children with higher vocabulary skill have greater familiarity with the events and roles we tested than children with lower vocabulary skill, due to increased exposure to spoken-language material, which conveys both new vocabulary and new events. There appears to be a tight relationship between vocabulary ability and amount of exposure to experiences that facilitate acquisition of early world knowledge such as parental speech and book reading (Hart & Risley, 1995; Hurtado et al., 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Raikes et al., 2006). Our measures of vocabulary ability may reflect relative differences between participants in their exposure to the event knowledge assessed by our study. Additionally, gross differences in age (children vs. adults) may covary with absolute differences in participants' event knowledge (adults have experienced far more events than children).

Our findings are consistent with both the processing ability and event familiarity accounts. When we control for age, we find associations between performance on our task and vocabulary ability. As described earlier, vocabulary ability may reflect differences in processing skill, in event experience, or both. Effects of participant age in our task suggest that overall world experience (tied to age) can also lead to improved performance on this task. These age effects support the notion that sheer differences in event familiarity, or general maturation, can facilitate real-time processing of talker-specific and verbal cues. Future work could refine the understanding of the nature of the relationship among processing ability, vocabulary skill and event knowledge via learning paradigms that can control current confounds among individual experience, vocabulary, and processing skill.

[1] The same pattern of age and vocabulary effects appeared when we included age as a continuous variable in our analysis.

Constraints on Theoretical Models of Sentence Comprehension

Importantly, our findings provide theoretical constraints for models of how listeners activate talker-relevant and event information as the sentence unfolds. In our study, we found that adults and children immediately generate inferences about the sentence's meaning using the talker's voice to activate world knowledge associated with that individual's role. Our participants did not rely solely on structural information at the action to guide interpretation or traditionally "linguistic" elements of the sentence as might be traditionally assumed by two-stage serial models of sentence interpretation (Frazier, 1995; Rayner, Carlson, & Frazier, 1983). Instead, our listeners took advantage of multiple sources of information, including a paralinguistic cue (talker-identity), from the earliest moments each source was available, consistent with constraint-based models of sentence interpretation (Elman, Hare, & McRae, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995).

The activated talker information is subsequently integrated with the action of the sentence to develop a strong expectation for a theme consistent with both talker and action cues. Locally relevant information is also activated to some smaller extent, as shown by the additional visual fixations toward the action-related item in the latter part of the sentence. This pattern is most consistent with sentence interpretation models in which listeners generate multiple expectations about very likely (globally coherent) information and less likely (locally coherent) information to occur as the sentence unfolds. Concretely, when a listener hears a pirate say, "I want to hold the . . .," not only does the listener develop an expectation for the globally coherent option, SWORD, but also considers a "holdable," locally relevant and less globally probable item, WAND. At first pass, this strategy may not seem conducive for efficient language interpretation. However, when we consider that much of real-world communication occurs under less than ideal listening environments, where talkers may misspeak or unexpected events may occur, it makes sense that listeners may employ sentence processing mechanisms that "keep their options open" as language unfolds. This local-coherence pattern echoes findings from other sentence interpretation studies (Kukona et al., 2011; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tabor, Galantucci, & Richardson, 2004) and bears a strong resemblance to rhyme effects in word recognition studies (Allopenna et al., 1998). Our findings add to a growing body of evidence that even very young children interpret language using mechanisms that allow for recovery in uncertainty (Borovsky et al., 2012; Fernald et al., 2010).

Limitations of the Current Study

We now note some remaining questions that suggest avenues for future work. A frequent question in developmental studies is to what extent task demands limit the performance of younger children. In our task, it is possible that the associations between the talkers and their roles were more challenging for younger participants to activate than for older children and adults. This concern is mitigated by the fact that even the youngest children showed qualitatively adult-like performance on this task. Minimally, this suggests that all of our participants successfully acquired and recalled the talker-identity associations, although there may have been developmental differences in the strength of this learned relationship between talker and identity.

An additional question is how much the results were driven by listeners' gender stereotypes. We always paired male and female voices, deliberately avoiding same-gender pairings, to ensure that all participants could effectively discriminate between the voice pairs in our task (for evidence that young children have difficulty discriminating same-gender voices, see Creel, 2012; Creel & Jiménez, 2012; Mann et al., 1979). However, we suspect that our results are not limited to gender-stereotyped pairs. Young child participants readily distinguish *familiar* same-gender talkers (Spence et al., 2002), such as their mother and other close female relatives. Further, research on preschool children's understanding of gender suggests that while gender is a salient category, children may make stronger inferences based on other characteristics, such as age (Taylor & Gelman, 1993) and race (Olson, Shutts, Kinzler, & Weisman, 2012). Thus, our results may indicate that children are activating specific roles (princess, cowgirl; pirate, astronaut) rather than generic gender biases. Whether our results reflect children's automatic activation of gender biases (women prefer wands and carriages) or specific role activation (*princesses* prefer wands and carriages), it is clear from our study that very young children readily use salient voice differences to activate pre-existing knowledge and can integrate that knowledge with sentential event structure.

In a similar vein, we presented our participants with talkers who had simple, unidimensional roles. Individuals in the real world are generally more complex, with multiple competing and occasionally contradictory facets to their personalities. How would adults and children associate a talker's voice with role cues that do not cohere with role stereotypes (e.g., a doctor speaking with a child's voice)? Would adults and children map multiple role features to a talker (e.g., could a talker be a pirate princess?) and interpret multiple potential preferences in real time? Future work would need to delineate how far these abilities extend in both children and adults.

Conclusion

Throughout the course of childhood, listeners develop an exquisite sensitivity to a variety of linguistic, paralinguistic, and nonlinguistic cues in the speech stream. This developing sensitivity must then be shaped via experience to allow listeners to weight these cues efficiently according to their informativeness during real-time speech comprehension. Our work adds talker identity to a growing list of cues that young children can use in real time to interpret speech. This is particularly interesting, given that talker information is sometimes regarded as an extraneous factor in language processing. While prior work has suggested that children fluently interpret linguistic cues in the speech signal, we have shown, for the first time, that children as young as ages 3–6 (our younger age group) make inferences about upcoming speech based on extralinguistic information about a talker's identity and without explicit mention of the talker's preferences for the items they talk about. To fully appreciate how children activate and interpret talker identity from voice cues alone, researchers in future work must investigate not only *when* this information becomes available to even younger listeners, but *how* listeners of all ages acquire and integrate this information in a rich and complex web of knowledge about individuals and their unique preferences.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, IL: Aldine.

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

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

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502–518. doi:10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583–609. doi:10.1111/j.1551-6709.2009.01022.x

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 22,* 527–565. doi:10.1080/01690960600845950

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology/Revue canadienne de psychologie, 27,* 464–472. doi:10.1037/h0082498

Boersma, P., & Weenink, D. (2012). Praat (Version 5.3.22) [Computer software]. Retrieved from http://www.praat.org/

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary and not age is associated with incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112,* 417–436. doi:10.1016/j.jecp.2012.01.005

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32,* 111–140. doi:10.1016/S0095-4470(03)00009-3

Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development, 83,* 2042–2056. doi:10.1111/j.1467-8624.2012.01816.x

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language & Linguistics Compass, 5,* 190–204. doi:10.1111/j.1749-818X.2011.00276.x

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology, 113,* 487–509. doi:10.1016/j.jecp.2012.07.007

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–534. doi:10.1080/01690960143000074

Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test*. Minneapolis, MN: NCS Pearson.

Elman, J. L., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature–nurture: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Erlbaum.

Experiment Builder [Computer software]. (2011). Mississauga, Ontario, Canada: SR Research.

Eyelink 2000 [Apparatus and software]. (2004). Mississauga, Ontario, Canada: SR Research.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83,* 203–222. doi:10.1111/j.1467-8624.2011.01692.x

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16,* 234–248. doi:10.1111/desc.12019

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42,* 98–116. doi:10.1037/0012-1649.42.1.98

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology, 60,* 190–217. doi:10.1016/j.cogpsych.2009.12.002

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. Fernandez, & H. Clahsen (Eds.), *Language processing in children* (pp. 97–135). Amsterdam, the Netherlands: Benjamins.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24,* 437–468. doi:10.1007/BF02143161

Hart, B., & Risley, T. (1995). *Meaningful differences in everyday parenting and intellectual development in young American children*. Baltimore, MD: Brookes.

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development, 12,* 213–238. doi:10.1016/S0885-2014(97)90014-9

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45,* 1723–1739. doi:10.1037/a0016704

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137,* 151–171. doi:10.1016/j.actpsy.2010.11.003

Huettig, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology, 2,* 285. doi:10.3389/fpsyg.2011.00285

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11,* F31–F39. doi:10.1111/j.1467-7687.2008.00768.x

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248. doi:10.1037/0012-1649.27.2.236

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156. doi:10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32,* 37–55. doi:10.1023/A:1021933015362

Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language, 32,* 855–876. doi:10.1017/S0305000905007051

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, USA, 104,* 12577–12580. doi:10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27,* 623–634. doi:10.1521/soco.2009.27.4.623

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38,* 618–625. doi:10.1016/S0022-1031(02)00510-3

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119,* 23–42. doi:10.1016/j.cognition.2010.12.002

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104. doi:10.1121/1.1908694

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18,* 193–198. doi:10.1111/j.1467-9280.2007.01871.x

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703. doi:10.1037/0033-295X.101.4.676

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 843–847. doi:10.1037/a0029284

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153–165. doi:10.1016/0022-0965(79)90067-5

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11,* F9–16. doi:10.1111/j.1467-7687.2008.00671.x

McAuliffe, M. J., Gibson, E. M., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *Journal of the Acoustical Society of America, 134,* 1358–1368. doi:10.1121/1.4812764

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5,* 1–10.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72,* 834–843. doi:10.1111/1467-8624.00318

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86,* 314–329. doi:10.1016/j.jecp.2003.09.001

Olson, K. R., Shutts, K., Kinzler, K. D., & Weisman, K. G. (2012). Children associate racial groups with wealth: Evidence from South Africa. *Child Development, 83,* 1884–1899. doi:10.1111/j.1467-8624.2012.01819.x

Peirce, C. S. (1998). A syllabus of certain topics of logic. In *The essential Peirce: Vol. 2. Selected philosophical writings, 1893–1913* (pp. 258–270). Bloomington: Indiana University Press (Original work published 1903).

Peterson, G. E., & Barney, H. L. (1952). Control methods use in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184. doi:10.1121/1.1906875

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

Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., . . . Rodriguez, E. T. (2006). Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development, 77,* 924–953. doi:10.1111/j.1467-8624.2006.00911.x

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358–374. doi:10.1016/S0022-5371(83)90236-0

Sedivy, J. C., Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147. doi:10.1016/S0010-0277(99)00025-6

Shafto, C. L., Geren, J. C., & Snedeker, J. (2010). Effects of maternal input on language in the absence of genetic confounds: vocabulary development in internationally adopted children. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society: August 11–14, 2010, Portland, Oregon* (pp. 2775–2780). Austin, TX: Cognitive Science Society.

Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49,* 238–299. doi:10.1016/j.cogpsych.2004.03.001

Spence, M. J., Rollins, P. R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language and Hearing Research, 45,* 214–222. doi:10.1044/1092-4388(2002/016)

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50,* 355–370. doi:10.1016/j.jml.2004.01.001

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C.(1995, June 16). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. doi:10.1126/science.7777863

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech, language, and communication* (pp. 217–262). San Diego, CA: Academic Press.

Taylor, M. G., & Gelman, S. A. (1993). Children's gender- and age-based categorization in similarity and induction tasks. *Social Development, 2,* 104–121. doi:10.1111/j.1467-9507.1993.tb00006.x

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134. doi:10.1016/S0010-0277(99)00032-3

Van Berkum, J. J. A., Van Den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20,* 580–591. doi:10.1162/jocn.2008.20054

(*Appendix follows*)

Appendix

All Quartets of Stimuli Used to Generate Sentences

Quartet	Talkers		Verbs		Talker 1 pictures		Talker 2 pictures	
1	Pirate	Princess	Hold	Ride in	SWORD	SHIP	WAND	CARRIAGE
2	Astronaut	Cowgirl	Ride on	Wear	SPACESHIP	SPACESUIT	HORSE	HAT
3	Child	Pilot	Drink	Fly	JUICE	KITE	COFFEE	PLANE
4	Soldier	Baby	Drink	Wear	WATER	BOOTS	MILK	DIAPER
5	Mom	Farmer	Drive	Feed	CAR	BABY	TRACTOR	PIG
6	Girl	Dad	Eat	Wear	CANDY	DRESS	HAMBURGER	TIE

Received June 2, 2013
Revision received August 23, 2013
Accepted October 1, 2013 ■