CHILD DEVELOPMENT

Child Development, xxxxx 2012, Volume 00, Number 0, Pages 1–15

Preschoolers' Use of Talker Information in On-Line Comprehension

Sarah C. Creel
University of California, San Diego

A crucial part of language development is learning how various social and contextual language-external factors constrain an utterance's meaning. This learning process is poorly understood. Five experiments addressed one hundred thirty-one 3- to 5-year-old children's use of one such socially relevant information source: talker characteristics. Participants learned 2 characters' favorite colors; then, those characters asked participants to select colored shapes, as eye movements were tracked. Results suggest that by preschool, children use voice characteristics predictively to constrain a talker's domain of reference, visually fixating the talker's preferred color shapes. Indicating flexibility, children used talker information when the talker made a request for herself but not when she made a request for the other character. Children's ease at using voice characteristics and possible developmental changes are discussed.

One of the greatest questions in development is how children begin to understand spoken language. We know that by adulthood, listeners use information at a variety of linguistic levels—phonology, syntax, discourse—and even extralinguistic information, such as visual scene cues (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) or affordances (Chambers, Tanenhaus, & Magnuson, 2004)—to reach full comprehension. However, it is not clear how readily children use extralinguistic information. In some cases, children seem to have difficulty integrating cues that adults use. It is not well understood how or when various cues are utilized during development, and what differentiates earlier acquired cues from later acquired ones.

One relatively underexplored cue that is potentially available to children is talker information. That is, children might enhance their understanding of spoken language by knowing what person, or what sort of person, is talking to them. One way to obtain this information is through the voice. Voices are somewhat like faces in their ability to convey both categorical information and individual identity (e.g., Mann, Diamond, & Carey, 1979; Perrachione, Chiao, & Wong, 2010). The speech signal contains abundant ''indexical information'' about the talker, including gender (Perry, Ohde, & Ashmead, 2001), age (e.g., Peterson & Barney, 1952), femininity (Ko, Judd, & Blair, 2006), sexual orientation (Munson, 2007; Pierrehumbert, Bent, Munson, Bradlow, & Bailey, 2004), region of origin (Bradlow & Bent, 2008; Clopper & Pisoni, 2004), socioeconomic status (Labov, 1966), and individual identity (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985). Any of these factors will be related to the content of that individual's language output, reducing the set possibilities of what that person is likely to say. As a very simple example, a female voice saying ''I need to go to the'' is more likely to continue with ''ladies' room'' than ''men's room,'' while the opposite is true of a male voice. Such talker-associated knowledge can potentially aid language comprehension.

What is not clear is whether children, like adults, use information about who is talking to inform their comprehension. Adults use talker information during on-line sentence comprehension (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008), but adults can use a number of cues in on-line processing that children cannot. Children are sensitive to acoustic variability in word-recognition and word-production tasks (Richtsmeier, Gerken, Goffman, & Hogan, 2009; Ryalls & Pisoni, 1997), suggesting that they may be sensitive to talker variability in comprehension as well. Below, I discuss how adults exploit talker information in language processing and then detail what is known

Thanks to Heather Pelton for running participants; to Carol Lee, Garrett Work, Shawn Walker, and Jordan Danly for providing voices; to Roger Levy for suggesting Experiment 5; and a big thank you to the children, parents, and preschool teachers and directors who made this research possible.

Correspondence concerning this article should be addressed to Sarah C. Creel, University of California, San Diego, Department of Cognitive Science, 9500 Gilman Drive Mail Code 0515, La Jolla, CA 92093-0515. Electronic mail may be sent to creel@cogsci.ucsd.edu.

 0009-3920/2012/xxxx-xxxx
DOI: 10.1111/j.1467-8624.2012.01816.x

about preschool children's use of talker information and other related characteristics of the speech signal.

Using Talker Information: Adults

Adult listeners can identify a large number of voices (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985), and they appear to relate voice characteristics to linguistic content (Church & Schacter, 1994; Geiselman & Crawley, 1983; Goldinger, 1996, 1998; Schacter & Church, 1992). For instance, Palmeri, Goldinger, and Pisoni (1993) presented listeners with a series of words for a lexical decision task. Each word was spoken by a particular talker. They then gave listeners a new series of words for an old–new judgment task. Of the ''old'' words, some were spoken by the original talker, and some were spoken by a new talker. Recognition as ''old'' was enhanced for words that were spoken by the original talker. Relatedly, Creel, Aslin, and Tanenhaus (2008) found that talker specificity aided on-line word recognition. Listeners repeatedly heard words spoken by particular talkers while they selected pictures of the words on a computer screen. After two phonologically similar words were repeatedly spoken by two different talkers—for example, a male voice saying ''sheet'' and a female voice saying ''sheep''—listeners made fewer visual fixations to the sheet when hearing a female voice beginning to speak ''sheep.'' In each case, recognition was facilitated by talker-specific match. These studies are consistent with the hypothesis that talker-specificity effects result from *precise acoustic encoding* of words that contains both phonemic information and voice characteristics (Goldinger, 1996, 1998).

Recent work shows that adults also make inferences based on talker identity during language processing. Van Berkum et al. (2008) showed that adults generate a larger semantic mismatch evoked potential to ''wine'' when they hear a sentence like ''I would like to have a glass of wine'' in a child's voice than when they hear it in an adult voice. This suggests that the expectedness of ''wine'' is modulated by inferred properties of the talker. Such inferences can have social consequences, when they influence judgments of competence (Ko et al., 2006) or masculinity (Pierrehumbert et al., 2004). These studies are consistent with the hypothesis (Geiselman & Crawley, 1983) that comprehenders encode talker information *semantically*: inferred properties of the talker influence processing and encoding of linguistic content.

This semantic-encoding account of talker-specificity effects contrasts with the acoustic-encoding account. In acoustic encoding, talker-specific advantages are thought to result from a more precise acoustic match between the current speech event and the previous (encoded) speech event. In semantic encoding, talker-specific advantages in processing stem from activation of knowledge about *people*, and voice characteristics are one way to activate this knowledge. In practice, it is difficult to pull apart semantic-encoding effects from acoustic-encoding effects, as most talker-specificity demonstrations are amenable to either explanation.

Using Talker Information: Children

Children seem to be exquisitely sensitive to acoustic information related to talker identity. Even before birth, fetuses encode information distinguishing their mother's voice from a stranger's voice (DeCasper & Fifer, 1980; Kisilevsky et al., 2003). During the 1st year of life, infants appear to have very acoustically specific representations of word forms: 7.5-month-olds have difficulty generalizing a newly familiarized word over gross changes in talker (Houston & Jusczyk, 2000), vocal emotion (Singh, White, & Morgan, 2008), and accent (Schmale & Seidl, 2009). When generalizing to new instances, infants benefit from high variability exposure (Rost & McMurray, 2009, 2010; Singh, 2008; see also Kovack-Lesh & Oakes, 2007, in visual categorization). Richtsmeier et al. (2009) found that older (preschool-aged) children produced words better when they heard the words with high talker variability. These studies suggest that children are sensitive to acoustic cues to identity from infancy, and that this acoustic sensitivity continues into the preschool years. This early sensitivity to acoustic form in the 1st year of life is consistent with an account of phonological development where young learners move from language-general to language-specific sound sensitivity, gradually focusing attention on information relevant to meaning in the target language (Werker & Tees, 1984). It is not clear that focusing on word-relevant information necessarily implies tuning *out* talker information as much as *not attending* to it; older children's (Richtsmeier et al. 2009) and adults' (e.g., Creel et al., 2008; Palmeri et al., 1993) sensitivity to talker information may be a remnant of the strong talker-specificity effects seen in very young children.

Thus, children are sensitive to speech acoustics, but can they use acoustic cues to identify individuals? To do this, children must both perceive acoustic cues and map them onto particular individuals (Mom) or groups (young female). There are some suggestions that children make such mappings. For instance, 5-year-olds state preferences to be friends with people who speak an acoustically familiar language (Kinzler, Dupoux, & Spelke, 2007). Hirschfeld and Gelman (1997) showed that children associate an unfamiliar language (Portuguese) with unfamiliar-looking clothing, dwellings, and people. This suggests that children can form associations between (un)familiar voice characteristics and (un)familiar people. Further, given children's early sensitivity to the acoustics of the speech signal, one might expect them to be expert at recognizing voices, but the evidence is somewhat equivocal. Children can identify the voices of classmates (Bartholomeus, 1973) and familiar cartoon characters (Spence, Rollins, & Jerger, 2002) with moderate accuracy. However, Mann et al. (1979) found that children ages 6–10 years were worse than adults at discriminating unfamiliar same-age same-gender voices in an XAB task; 6-year-olds were at chance. This result suggests that children's processing of unfamiliar voices may be *poorer* than adults,' perhaps reflecting a need for extensive perceptual learning or greater attentional capacity in order to adequately process acoustically-subtle voice distinctions.

Of course, even being able to identify a talker does not mean that children can use this information rapidly during language comprehension. Five- to six-year-old children often seem to process cues more slowly than adults, or fail to integrate cues with linguistic information. For instance, 5-year-olds have difficulty using visual scene information to reroute garden-path sentences (Trueswell, Sekerina, Hill, & Logrip, 1999), with adult-like performance not evident until age 8 (Weighall, 2008). Additionally, children use paralinguistic information—prosodic patterns and vocal-emotional information—less adeptly than adults. Ito and colleagues (Bibyk, Ito, Wagner, & Speer, 2009; Ito, Jincho, Yamane, Minai, & Mazuka, 2009) show that English-speaking and Japanese-speaking 6-year-olds are sensitive to contrastive prosodic patterns, but are slowed relative to adults in using prosody in real-time comprehension (see also Snedeker & Yuan, 2008). Children do not seem to apprehend the meanings of vocal emotional cues until age 4 or 5 (Quam & Swingley, 2012), may not use it predictively until after age 4 (Berman, Chambers, & Graham, 2010), and may disregard vocal emotional cues when they conflict with sentence content up to age 8 or so (Morton & Trehub, 2001). This suggests that preschool children's associations of prosody with meaning are present but somewhat fragile. Thus, preschool children may also have difficulty using talker information to constrain comprehension, despite recognizing a particular talker.

The picture painted by previous research leads to a number of possibilities about how preschool children might use talker information in on-line comprehension. First, they may not use it at all because it requires integrating talker information with other (sentential) sources of information, and children generally have more trouble than adults in integrating cues (as in Morton & Trehub, 2001; Trueswell et al., 1999; see also Zelazo, Frye, & Rapus, 1996). Second, children may show sensitivity to talker information in terms of low level acoustic matching (''I have heard this set of acoustic characteristics before'') but may not be able to map sound patterns onto higher level representations, such as particular individuals (''This sounds like the person who likes pink things''). This means that if children do use talker information to constrain sentence processing, they might be using it implicitly—showing a processing benefit due to greater low-level *acoustic* similarity to past experiences rather than accessing representations of the talkers themselves. Finally, children may be limited relative to adults in using talker information in comprehension, due to poorer abilities to distinguish novel voices (Mann et al., 1979). That is, they may be able to use highly distinct voices to constrain on-line comprehension but may not be able to use very similar voices.

The Present Study

Adults use voice cues to talker identity to constrain comprehension of spoken language, but children might not do so for a variety of reasons. Thus, the current study had three aims. The first was to investigate whether children use voice cues *at all* to talker identity to constrain sentence processing. Second, if children use voice cues, how do they use them—at a low (acoustic) level or a high (semantic) level? While adults presumably use voice cues to activate knowledge about the talker, children might simply associate acoustic cues with some external referent. The third aim was to explore whether children are limited in their abilities to use voice information due to difficulty discriminating voices.

These questions were addressed in a series of ''visual world'' language comprehension experiments (Cooper, 1974; Tanenhaus et al., 1995). In each, participants were instructed that each of two characters preferred a different color. They were then asked by each talker to select objects. The question was whether they would visually fixate each talker's preferred color, prior to knowing what object the talker was going to ask for. Experiments 1 and 3 asked whether children used talker information in sentence processing at all, measuring whether they fixated shapes in each talker's preferred color, using gender-stereotyped colors for maximal ease in Experiment 1, and novel, non-gender-stereotyped talker-color mappings in Experiment 3. Experiment 2 considered whether participants fixated shapes based on a low-level linkage of acoustic attributes to color attributes, or based on semantic knowledge of particular individuals (the characters themselves). Experiments 4 and 5 explored whether or not children were limited in their abilities to use more similar voices (and more similar characters) to inform on-line comprehension.

Experiment 1

The foremost question in this study is whether children are at all able to use acoustic information linked to the talker to guide comprehension of spoken language. Therefore, Experiment 1 considered whether, in a very simple situation, children could use knowledge about each of two characters to interpret identical spoken sentences differently. Children learned the color preferences of two characters—one male, one female—and then aided the characters in selecting shapes of their preferred colors. To make the task as easy as possible, the favorite colors corresponded to common gender stereotypes. Importantly, children could perform the task—selecting shapes—with perfect accuracy without invoking knowledge about each character. However, if they routinely invoke knowledge about the person speaking when processing a sentence, then visual fixations should be biased toward the talker's preferred color.

Method

Participants. $N = 24$ English-speaking children (6 female), ages 3–5 years ($M = 4.23$, $SD = .44$), and $n = 24$ English-speaking adult controls took part. Three more children were tested but excluded due to low accuracy (2) or excessive data loss (> 20% of samples with no gaze information; 1). Children were recruited from and tested at day cares and preschools in the San Diego area. Adults were recruited from the University of California San Diego participant pool, and were in the lab for other, unrelated experiments. The sample was predominantly White, with East Asian, South Asian, and African American ethnicities represented. Data loss by the eye tracker—that is, inability to detect gaze location, for instance, due to blinking or excessive child movement—was not large for children ($M = 7.5\%$, $SD = 5.4\%$) or for adults ($M = 6.1\%$, $SD = 4.9\%$).

Stimuli. Visual shape stimuli were generated in PowerPoint and exported as 200×200 pixel jpg files. Anna and Billy characters were created using the South Park Avatar Creator (http://www.southparkstudios.com/avatar/), and a scene of each character surrounded by objects of their preferred color was created in image-editing software.

Auditory stimuli were recorded by two English-speaking Southern California natives (CL and GW; see online supporting information Appendix S1 for voice characteristics). The stimuli in Table 2 were recorded in a sound-attenuated booth on a high-quality microphone and were edited digitally to remove silences and set volumes to 70 dB SPL.

Equipment. In all experiments, visual fixations were recorded by an Eyelink 1000 Remote eye tracker (SR Research, Mississauga, ON, Canada; http://www.sr-research.com). This instrument has 2-ms temporal precision and spatial precision of 0.5° visual angle. It requires no equipment to be worn by the participant, only a small sticker, allowing easy and comfortable use with young children as well as adults. The eye tracker was run by a Dell tower in DOS mode. Stimuli were presented by a Mac Mini running Matlab, from scripts using PsychToolbox 3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002).

Procedure. Prior to the actual experiment, children's color preference data were recorded to verify that the population studied did in fact demonstrate the color preferences intended to be gender-stereotyped. These data were collected for Experiments 1–4, and are presented in Table 1. Girls most often named pink (44.4% of girls), and boys, blue (34.1%).

After an eye tracker calibration sequence, the experiment proper began. It had four phases (Table 2). In the *introduction*, each talker appeared surrounded by objects of their favorite color, and stated a preference for that color. In the *color check*, children saw two shapes that were identical except for color, and one of the characters asked the child to find the ''[preferred-color] one.'' This verified that each child could distinguish the colors, and reinforced children's knowledge of color preferences. Then, the two introduction trials repeated. Finally,

Table 1
Favorite Colors, Experiments 1–4

Color	Girls %	Boys %
Blue	3.3	**34.1**
Pink	**44.4**	2.2
Red	5.6	24.6
Purple	21.1	2.2
Green	2.2	10.9
Yellow	4.4	6.0
Black	1.1	6.5
Orange	0.0	3.0
Brown	0.0	1.5
Gold	0.0	1.5
No favorite	0.0	1.5
Rainbow	2.2	0.0
Turquoise	2.2	0.0
Missing data	13.3	6.0

Note. Children who specified multiple (*n*) favorite colors were counted as one *n*th of a child for each color. The most selected color for each gender is bolded.

Table 2
Order of Trials in Each Experiment

Trial type	No. of trials	Material
Introduction[a]	2	Hi, I'm Anna. I like things that are pink. I like my pink bed, and my pink tutu, and my pink bunny slippers. Can you help me find pink things?
Color verification[a]	≥8	Anna: Where's the pink one? …
Repeat of introduction[a]	2	Hi, I'm Anna. I like things that are pink ….
Test trials[b] Experiments 1 3 and 4	32	Anna: Can you help me find the square? Billy: Can you help me find the circle?
Test trials, Experiment 2	32	Anna: I want to see the square. Can you show me where it is? Billy: Anna wants to see the circle. Can you show her where it is?
Test trials, Experiment 5	32	Anna wants to see the square. Can you show her where it is? Becky wants to see the circle. Can you show her where it is?

[a]In Experiments 3–5, pink and blue became white and black or black and white. In Experiment 5, introductions and color verification were in the third person.
[b]In Experiments 4–5, Anna and Billy became either Anna and Becky or Adam and Billy.

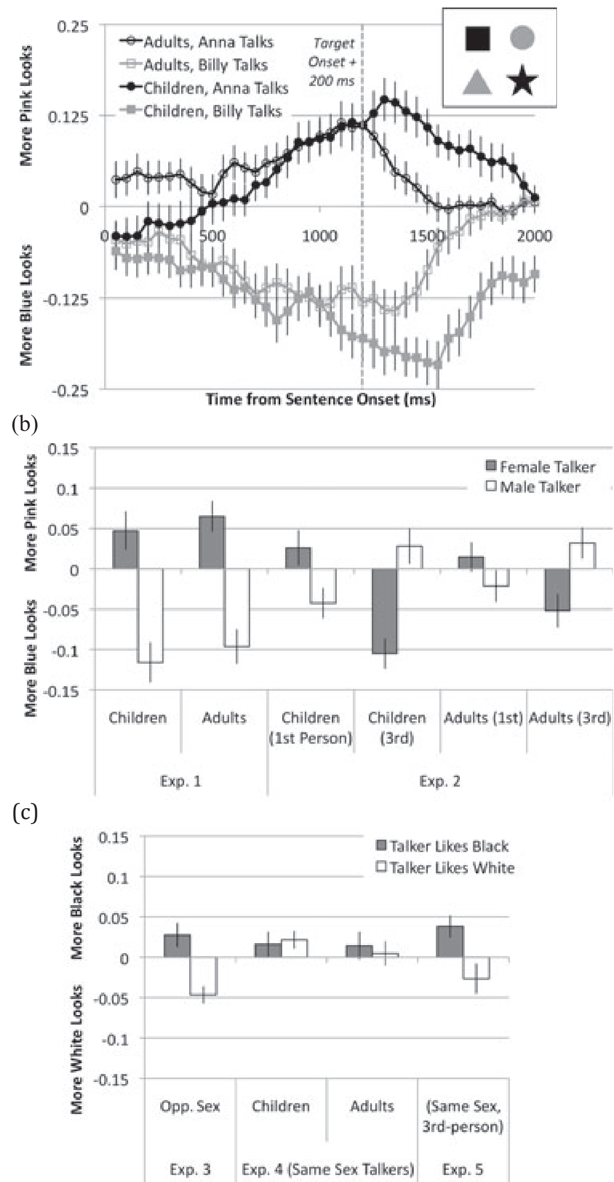

Figure 1. (a) Experiment 1, looks to pink shapes minus looks to blue shapes on Anna trials and Billy trials, with standard errors, for child and adult participants. A score of zero represents equal likelihood of fixating a particular color. Inset: a sample shape display (black = pink; gray = blue). (b) Average looks (with standard errors) to pink shapes minus blue shapes, Experiments 1 and 2, from 200 to 1200 ms. (c) Average looks (with standard errors) to black shapes minus white shapes, Experiments 3–5, from 200 to 1200 ms.

participants received 32 *test trials* (Figure 1a, inset). On each, they saw four shapes—two blue and two pink—and heard one of the two characters say, ''Can you help me find the X?'' X was equally likely to be a square, circle, triangle, or star. The color of the shape was never mentioned on the test trials, nor did the character speaking appear onscreen during the test trials. This means that children had to recognize each character from her or his voice. The talker was a perfect predictor of the *color* of the target shape (e.g., when Anna spoke, the target was always a pink shape)—but not of the target shape itself (there was always another shape on screen that had the same color as the target). Thus, children could never predict the identity of the target from the talker. Children indicated responses by pointing, and an experimenter clicked the computer mouse on that picture to record the response. Adult participants clicked pictures with a mouse on their own.

Results

Looks to blue nontarget shape(s) were subtracted from looks to pink non-target shape(s) to form a pink-preferring looking score. When there were two nontarget shapes of the same color in a display (e.g., two blue shapes), their looking proportions were averaged. If listeners are sensitive to talker gender, they should look more to pink shapes when the talker is female than when the talker is male. Prior to subtraction, looking proportions underwent an empirical logit transformation (Barr, 2008) to correct for non-normality in proportions. The window for analysis of looking times was 200–1200 ms. This window was selected because it roughly spanned the time point between the onset of the sentence to the time point where the shape word began (marked on each figure). A 200-ms delay on these time points (0 + 200 ms, 1000 + 200 ms) is standardly included in eye tracking analyses as the time to plan and execute an eye movement based on an external signal (Hallett, 1986). For consistency, this time window was maintained for all following experiments. Throughout, target looks are omitted from the pink-preference score so that any early looks based on the shape name are excluded. This isolates out visual fixation patterns based on color-preference knowledge alone.

Careful examination of Figure 1a, which depicts the pink-looking bias on Anna trials and Billy trials, suggests that children are using talker information to aid them in sentence comprehension. Specifically, there are more looks to pink things early on when Anna is talking, and fewer looks to pink things (i.e., more looks to blue things) when Billy is talking. Adult listeners (hollow shapes in Figure 1a) showed a similar pattern to the children, though they seemed to resolve ambiguity faster than the children. This pattern of faster adult processing, present throughout, has been noted by other researchers (Sekerina & Brooks, 2007). Error rates were low in all experiments and mostly unrelated to color. Because errors were uninformative they are not further discussed in the text but are reported in Appendix A for completeness.

For children, an analysis of variance (ANOVA) was conducted on the mean difference in empirical-logit-transformed looking times to pink shapes minus blue shapes in the 200–1200 ms time window (pink preference score; Figure 1b, left) with talker gender (female, male) as a within-participants factor, confirming the above observations. There was an effect of talker gender, $F(1, 23) = 14.57$, $p = .0009$, with higher pink preference on female-talker trials than male-talker trials. Interestingly, even though the age range was fairly broad, a correlation of age with the preferred–nonpreferred color difference score for each participant was not significant in this ($r = -.22$), $t(20) = 1.01$, $p = .33$, or other experiments (Appendix B). This may be the case because even the younger children are able to use talker information. An ANOVA for adults showed a similar pattern to children: An effect of talker gender, $F(1, 23) = 18.71$, $p = .0003$, reflected more pink looks on female-talker trials than male-talker trials. Similar analyses with experiment half (first 16 trials, second 16 trials) as a factor indicated no effects of or interactions with experiment half in any experiment ($F \leq 1.31$, $p \geq .26$; all but two Fs < 1), suggesting that no detectable learning occurred during the test phase. For space reasons, analyses in the text collapse over this factor.

Discussion

Preschool-aged children and adults rapidly executed eye movements to shapes of the talker's preferred color. This suggests that children, as well as adults, can use acoustic cues to talker identity to activate knowledge about the person speaking and then use that knowledge to constrain the domain of reference for that talker.

However, there are at least two plausible explanations for children's performance on this task. Adults can presumably use talker information to make inferences about a talker's likely behavior (Van Berkum et al., 2008). For children it is less clear what the case is. They might be making—or might already possess, given the choice of gender-stereotyped colors—low-level mappings between color and the vocal properties of the talker. This would predict that merely hearing Anna speak would activate the color pink and cause more pink-shape looks, regardless of the content of what she said. However, this low-level mapping would not be particularly useful, because if Anna were talking about someone besides herself, Anna's color preference would be irrelevant. On the other hand, children may be using talker information at a high level—to determine whose preferences should be invoked in constraining processing. That is, they might know to utilize acoustic cues selectively, when the talker is speaking about herself.

To distinguish between these two interpretations, Experiment 2 considered what happens when talker is only sometimes informative. Half the time, each character asked for a shape for herself, and the other half of the time, the character asked for the other character. If participants simply make low-level auditory-visual associations, then they should look at the talker's preferred color even when the talker refers to the other character. If, instead, participants are using talker-related acoustic variation to access information about the individuals involved in the sentence, then they should look to the talker's preferred color on first-person trials, and to the talker's nonpreferred color (but the *agent of the sentence's* preferred color) on third-person trials.

Experiment 2

Method

Participants. $N = 32$ adults and $n = 31$ children, aged 3–5 years (15 female; $M = 4.26$, $SD = .54$), were drawn from the same sources as in Experiment 1. N was increased slightly due to the fact that the number of trials per condition per participant was halved relative to Experiment 1. Three more children's and three more adults' data were excluded due to poor eye track (> 20% data loss), and two children were excluded for low accuracy. Data loss in the final sample was relatively small (children: $M = 7.5\%$, $SD = 5.2\%$; adults: $M = 5.4\%$, $SD = 4.6\%$).

Stimuli. The same visual stimuli as in Experiment 1 were used. The same talkers as in Experiment 1 recorded new sound files for the test phase, as in (1) and (2) below. The syntactic form of the instruction was changed so that the wanter's name was early in the sentence on third-person trials; that is, the wanter was the agent of the sentence, ensuring that listeners knew both who the wanter was and who the talker was from the same point in time.

1. I want to see the square. Can you show me where it is?
2. Anna [Billy] wants to see the square. Can you show her [him] where it is?

Procedure. This matched Experiment 1, except that in the test phase, 16 trials were first-person trials (''I want to see'') and 16 were third-person trials (''Anna wants to see'').

Results

In this and following experiments, data are depicted in terms of pink preference—looks to pink shapes minus looks to blue shapes (Figure 1b; this changes to black shapes minus white shapes in Experiments 3–5, in Figure 1c). In the current experiment (Figure 1b, center), when Anna asked for a shape for herself, pink preference was positive, but when Anna asked for a shape for Billy, pink preference was negative (i.e., they looked more at blue things). That is, the pink preference effect reversed from the first-person condition to the third-person condition. Adults (Figure 1b, right) performed similarly, though effects on first-person trials were small.

ANOVAs on pink preference with person (first, third) and talker gender (female, male) substantiated these observations. For children, no main effects reached significance, but there was an interaction of Person × Talker Gender, $F(1, 29) = 17.42$, $p = .0002$, indicating that there were more pink looks on female-talker trials when sentences were in the first person, but more pink looks on *male*-talker trials (i.e., female-*agent* trials) when sentences were in the third person. Talker gender was significant on first-person trials, $F(1, 29) = 4.51$, $p = .04$, and was significant in the opposite direction on third-person trials, $F(1, 29) = 18.77$, $p = .0002$.

For adults, no effects were significant, though there was a marginal interaction of Person × Talker Gender, $F(1, 31) = 3.57$, $p = .07$. Looking more closely, the preferred-color looking effect was significant only on third-person trials, $F(1, 31) = 6.50$, $p = .02$; first-person, $F(1, 31) < 1$, $p = .52$. The overall looking pattern was similar to children's. It is not clear why adults would show no first-person effect when they did so in Experiment 1. It is worth noting that adults' looks trended toward significance over the course of the time window ($p = .053$ in the second half of the time window), though the effect magnitude is much diminished from Experiment 1. A distinct possibility is that adults have learned through experience that sentential content is a more reliable cue than talker identity, and when these cues are put in conflict, adults allocate more attention to the a priori more reliable cue.

Discussion

The eye tracking data suggested that even child participants were accessing high-level representations of individuals rather than merely activating a direct auditory-visual association between acoustic attributes and color. Children visually fixated the shapes that were the *agent's* preferred color, even if that color did not match the talker's preferences. When Anna asked for a shape for herself, as in Experiment 1, children looked more at the pink shapes. Crucially, when Anna asked for a shape on Billy's behalf, participants looked more at the blue shapes. This suggests that when children use talker information in on-line processing, they are using it to access information about the individual rather than merely activating representations of color based on acoustic information. Moreover, they can use it adeptly, switching between referential (third-person) cues and acoustic information to make predictions.

Thus, children appear to use voice characteristics to access their knowledge about people, which they then use to augment their comprehension. This capacity might be useful to them in everyday social interactions. However, before supposing that children routinely use talker information in language processing, it is important to know how much children's performance in this task is buttressed by long-term knowledge of gender-stereotyped color preferences. That is, children may form representations of the characteristics of individuals, or groups of individuals—such as genders—very gradually. On the other hand, children might be able to make talker mappings fairly easily, which would be more useful for them in interacting with new individuals. This was explored in the next experiment, which assessed whether children could learn previously unfamiliar color preferences for the two talkers, using non-gender-stereotyped colors—black and white—as the preferred colors instead of pink and blue.

Experiment 3

Method

Participants. $N = 24$ children, aged 3–5 years (8 female, 1 unspecified; $M = 4.54$, $SD = .57$), from the same pool as in previous experiments took part. Five additional children took part but were replaced due to low accuracy (2), failure to complete the experiment (2), or excessive track loss (1; > 20%). Adults were not tested, as it was assumed that they would be able to learn this mapping readily. Data loss in the final sample was relatively small ($M = 7.3\%$, $SD = 5.2\%$).

Stimuli. New shapes were created in black and white, with a mid-gray background. Note that the data in Table 1 support the idea that black and white are not generally preferred by either gender, though a small number of boys reported black as a favorite color. New Anna and Billy scenes were also created, with black- and white-preferring versions. These resembled the pink and blue scenes in Experiments 1 and 2, except that Billy's watergun was replaced with a black or white soccer ball due to the resemblance between a black watergun and a real handgun.

Procedure. This matched Experiment 1, except that pink and blue were replaced by black and white. Color preference was counterbalanced: For half of the participants, Anna preferred black and Billy white, and for the rest, Anna preferred white and Billy black.

Results

Fixations (Figure 1c, left) largely mirrored Experiment 1: Children looked more to the preferred-color shape of the individual speaking. An ANOVA on black preference with colors assigned (Anna = black, Billy = white, or the reverse) as a between-participants factor and talker color (preferred black, preferred white) as a within-participants factor showed an effect of talker color, $F(1, 22) = 8.11$, $p = .009$, with more looks to black shapes when the talker preferred black than when the talker preferred white. No other effects approached significance.

Discussion

These results demonstrate that children rapidly learn to ascribe information (color preferences) to particular vocally identifiable individuals. More generally, the results suggest that use of acoustic talker cues to constrain sentence processing can be applied to relatively recently learned information rather than requiring lengthy environmental exposure. This means that children may be able to use talker information to constrain language processing in a variety of new contexts, to the extent that they understand what a talker's particular predilections are.

Another variable that may affect the usefulness of talker cues in children's language processing is the level at which children can make distinctions between voices. They may be able to use acoustic cues to *individuals*. However, the individual talkers in Experiments 1–3 differed not only in identity but also on a very salient social dimension—gender. Thus, the results up to this point do not distinguish between children learning and using information about individuals versus learning and using information about social categories. To resolve what children (and adults) were learning, Experiment 4 taught participants the color preferences of two same-gender talkers, either Anna and Becky or Adam and Billy. If children are learning about the color preferences of individuals, then they should be able to distinguish the two characters' preferences. On the other hand, if children are using representations of genders rather than individuals, they will not be able to distinguish the same-gender characters' preferences.

Experiment 4

Method

Participants. $N = 30$ children (13 female; M = 4.69, SD = .52) and $N = 16$ adults from the same populations as in previous experiments took part. Five additional children's data were excluded for low accuracy (2), failure to complete the experiment (1), or data loss over 20% (2). Data loss was relatively small (children: M = 6.3%, SD = 4.5%; adults: M = 4.5%, SD = 3.5%).

Stimuli. Two additional characters, Adam and Becky, were created. Two additional natives of Southern California (SW and JD) recorded Adam and Becky stating preferences for white and black things, and requesting circles, squares, triangles, and stars. Each participant heard speech from either two female talkers (Anna, Becky) or two male talkers (Adam, Billy). Female speakers differed in second-formant frequency (F2) and in first-formant frequency; male speakers differed in fundamental frequency, F2, and speech rate (details in online supporting information Appendix S1).

Procedure. This matched Experiment 3, except that listeners heard same-gender pairs of characters. One fourth of participants each learned that: Anna liked white, and Becky, black; Anna, black and Becky, white; Adam, white and Billy, black; Adam, black and Billy, white.

Results

Neither children nor adult participants (Figure 1c, center) fixated preferred-color pictures based on talker information. For children, an ANOVA on black preference with talker gender and talker color as within-participants factors showed no main effects (Fs < 1). A marginal interaction of Talker Gender × Talker Color, $F(1, 28) = 3.21$, $p = .08$, resulted from a nonsignificant tendency on male-talker trials to look to the *non*preferred color, $F(1, 13) = 2.66$, $p = .13$, and a nonsignificant tendency ($F < 1$) on female-talker trials in the opposite direction. Adult participants showed no effects or interactions (Fs < 1).

Discussion

Unlike the preceding experiments, children and adults showed no evidence of using talker acoustics to invoke different color preferences. For adults, this is consistent with reports of ''change deafness'' (Vitevitch, 2003; see also Creel & Tumlin, 2011)—failure to notice a change in talker, similar to change blindness effects in the visual domain (Simons & Levin, 1998). In fact, when we conducted Experiment 4 with adults who knew ahead of time that they should attend to voice information, they readily made anticipatory looks to the characters' favorite colors, suggesting that both voice pairs were discriminable. Rather than an inability to discriminate voices, it implies that adults may not *attend* to differences between unfamiliar voices that are not extreme.

Why did children fail to use within-gender voice differences to make predictions in on-line language processing? It is possible that children, like the naïve adults, have already learned not to attend to voice characteristics that are not indicative of social differences (e.g., gender) between talkers. Another explanation is that when children learn information about a talker's preferences, they ascribe preferences to social groups (such as gender) rather than to individuals. That is, they can easily distinguish two same-gender voices but do not consider the possibility that two similar cartoon individuals of the same perceived gender (and age and social class and paraphernalia) would differ in their color preferences. Finally, it is possible that children, unlike adults, were unable to distinguish the two talkers from each other. This is consistent with the results of Mann et al. (1979): The youngest children tested (age 6) were at chance in detecting a same-gender voice change (see also Bartholomeus, 1973, who found that children were worse at identifying classmates' voices than their [adult] teachers were). The children in the current study were even younger, and thus it is plausible that they would find distinguishing unfamiliar voices difficult, even with exposure to the voices in a relatively naturalistic task.

A final experiment attempts to rule out the explanation that children fail to use voice cues in Experiment 4 due to difficulty representing similar characters. Much like the third-person condition of Experiment 2, Experiment 5 replicates Experiment 4, but with the demand for voice memory removed: All information is presented by a single third-person narrator. That way, children have a referential cue to the character rather than a voice cue. If the sole difficulty for children in Experiment 4 was distinguishing two relatively unfamiliar voices, then they should be able to guess ahead about what a named character will request, showing anticipatory looks to the character's preferred color. If, however, children have additional difficulty in encoding that two same-gender characters had differing preferences, then they should not show anticipatory looks to the characters' preferred colors.

Experiment 5

Method

Participants. $N = 21$ children, aged 3–5 years (11 female; $M = 4.71$, $SD = .72$), from the same pool as Experiments 1–4 took part. Eight additional children participated but were excluded due to low accuracy (4), failure to complete the experiment (1), or track loss greater than 20% (3). Data loss in the final sample was relatively small ($M = 10.6\%$, $SD = 5.5\%$).

Stimuli. Visual stimuli were the same as in Experiment 4. New sound files were recorded where each character was described by a single narrator in all phases of the experiment: ''This is Anna. Anna likes things . . .'' Similar passages were recorded for Becky, Adam, and Billy. Anna and Becky phrases were recorded by a female talker (HP), and Adam and Billy phrases were recorded by a male talker (CC). A given child only heard one narrator for the entire experiment.

Procedure. This matched Experiment 4, except that auditory materials from that experiment were substituted with new materials that described characters in the third person.

Results

As evident in Figure 1c (right), children used third-person reference to characters to make looks to preferred-color pictures. An ANOVA on black preference with talker color as a within-participants factor and talker gender as a between-participants factor showed an effect of talker color, $F(1, 19) = 6.99$, $p = .02$, suggesting that children looked more at black shapes when the talker preferred black things. An effect of talker gender ($F(1, 19) = 8.32$, $p = .009$) suggested more looks to black shapes overall for female talkers than male talkers, but this did not interact with talker color ($F < 1$).

Discussion

In this final experiment, children were able to ascribe different color preferences to same-gender characters. That is, they learned and used this preference information on-line to constrain their processing of characters' (third-person) requests. This rules out the hypothesis that that children in Experiment 4 were unable to represent differing preferences within a gender. However, it is consistent with the more probable hypothesis that those children failed to distinguish two same-gender voices. Of course, it is still possible that preschool children, like adults, have already learned to ignore voice changes when listening for content. Distinguishing these possibilities—inability to discriminate versus ignoring talkers—would be best approached by testing voice identification directly and parametrically manipulating the characteristics that differ between the to-be-distinguished voices, an undertaking beyond the scope of this article.

General Discussion

Five experiments demonstrated that preschool-aged children, as well as adults, use acoustic cues to who is talking to constrain on-line language processing. In Experiment 1, children and adults learned that each of two talkers preferred a gender-stereotyped color (pink or blue). Upon hearing that talker, they directed visual fixations to shapes of that talker's preferred color. In Experiment 2, children and adults learned the same contingencies but heard each talker request shapes equally often for herself or for the other talker. This meant that talker-related acoustics alone were not a valid cue to target color. Here, children looked to the *sentential agent's* preferred color, using talker acoustics only when agents were not explicitly named.

The next three experiments fleshed out the knowledge that child and adult listeners glean from talker-related acoustic cues. Experiment 3 questioned whether rapid, real-time use of talker-specific knowledge developed only over a lengthy period of learning, or could instead be calculated based on brief experience. With exposure to new color-preference contingencies (black vs. white), children again visually fixated the talker's preferred color. Experiment 4 showed that children, like adults, did not use talker information in on-line comprehension when talkers matched in gender, age, and dialect. Ruling out the hypothesis that children had trouble representing two similar characters, Experiment 5 presented all information about same-gender characters in the third person (similar to the third-person condition of Experiment 2), so that children had a referential cue rather than acoustic cues to identify characters. Here, children successfully used third-person reference to same-gender characters to constrain processing.

Returning to the original questions, human listeners are able to use acoustic cues to talker to constrain on-line language processing as early as the preschool years. Preschoolers appear to accomplish this by using voice characteristics to access representations of talkers' (or agents') mental states, specifically, their color preferences. Preschoolers rapidly learn and use simple preference knowledge about new talkers. However, neither preschoolers nor adults take advantage of vocal cues to identify two newly learned characters matched for gender, age, and dialect. For adults, this likely results from inattention rather than inability to discriminate the two voices. In children's case, difficulties are likely due to failure to discriminate, which itself may be due to inattention to talker characteristics.

Children's Use of Paralanguage in On-Line Processing

A relevant question is how preschool children's use of talker identity for comprehension relates to their use of other paralinguistic cues in the speech signal, such as vocal emotion and various prosodic phenomena (pitch accent, pauses). Preschoolers use emotional and prosodic cues in on-line comprehension to some extent. For prosody, Snedeker and Yuan (2008) found that 4- to 6-year-old children used phrase-final lengthening to resolve global ambiguity, and Ito and colleagues (Bibyk et al., 2009; see also Ito et al., 2009, for similar patterns in Japanese) found that children ages 6 and up were able to use pitch accent in comprehension. In all cases, children used prosodic cues more slowly than adults. A similar pattern of results shows up for vocal emotion. Berman et al. (2010) showed that 4-year-olds used vocal cues to emotion to interpret sentences. However, Berman et al.'s pre-target-word visual fixation patterns were overwhelmed by looking biases toward ''sad'' objects (e.g., a broken doll), so that emotionally congruent looks did not show up until the target word itself. Thus, we do not know if those children used vocal emotion cues predictively. Overall, compared to data on prosody and vocal emotion usage, children seem somewhat more advanced in using talker information in that they use it rapidly—prior to target word onset, and with comparable speed to adults—and alternate flexibly between using and not using talker information to interpret language (Experiment 2). However, it should be kept in mind that the current task is simpler than those used in the studies just discussed, meaning that children in a more difficult task might—as with prosodic and vocal-emotional information—be slower than adults to utilize talker information. It should also be kept in mind that voices are distinguished by a number of factors, including prosodic characteristics, but also phonetic characteristics (such as formant frequencies and frication noise; Newman, Clouse, & Burnham, 2001; Singh & Murry, 1978), and as such may be more recognizable.

Assuming for the moment that children do use talker information more robustly in comparison to other types of paralinguistic information, an obvious question is why. This might result from greater stability of the acoustic cues themselves or from greater stability of the associated mental-state attributes that the acoustic patterns map onto. Information distinguishing the gender of a talker may be more acoustically reliable than cues to vocal emotion. It is important to note here that an acoustically *salient* cue is not necessarily acoustically *reliable*. For example, certain sound properties related to vocal emotion may be very acoustically salient, such as shouting when angry. However, shouting may be a low-reliability cue—someone may shout for reasons other than anger (to be heard above noise, or due to a hearing impairment or poor personal volume management)—and some angry people may not shout. In contrast, there are not many reasons for an adult to have extremely low fundamental and formant frequencies besides being male. Thus, children may have more difficulty mapping vocal emotional displays to affective states than in mapping talker-related acoustic information to talker's preferences.

Another potential advantage for learning voice-mental attribute associations is that more stable mental-state information can be linked to talker cues but not to vocal-emotional cues. That is, cues to vocal emotion indicate (usually) transient mental state information about an individual, while cues to talker indicate somewhat more stable mental state information. A better understanding of the development of paralinguistic cue use is likely to come from careful examination of the reliability of both the acoustic structures themselves and the concepts with which they are associated.

Possible Limitations in Use of Talker Identity

It should be noted that it is not clear whether preschool children are able to use talker-related acoustic information in a particularly nuanced way, for instance, at the level of the individual. The evidence presented is consistent with the notion that children learn information about the preferences of individuals but can only use fairly large acoustic differences to make predictions during sentence processing. Nonetheless, we know that preschool children can identify familiar voices (Bartholomeus, 1973; Spence et al., 2002). It stands to reason that provided children can individuate a particular voice, they should be able to access information about the individual with which it is associated. For instance, a child with two sisters might be able to use those two familiar young female voices to make predictions about how they may continue a sentence. It is also possible that more acoustically distinct same-gender voices might yield positive results. Note, though, that in similarity-scaling studies, gender often accounts for the largest amount of variance (e.g., Singh & Murry, 1978), suggesting that any gender difference is likely to be more salient than a within-gender voice difference.

The limitations of the current task, too, must be acknowledged: Clearly, there is much more to learn about individuals than their favorite colors. To what extent are children able to use voices to access more nuanced information about talkers? Defining children's abilities in this regard may be dependent on establishing how much information mature (adult) comprehenders are able to glean from voice information. As yet, the work by Van Berkum et al. (2008), where adult listeners use nuanced talker information to interpret sentences, is somewhat unique, meaning that more investigation is needed of adult abilities to use talker-related knowledge in comprehension. Nonetheless, two factors are likely to affect children's (and adults') use of voice information: how well encoded the voice-individual mapping is in memory, and how well encoded information about the individual is in memory. The current work suggests that the voice-individual mapping may provide difficulty for children, perhaps due to weaker auditory representations of voice variability, or due to their more limited memory resources being allocated predominantly to word learning (e.g., Newman, 2008). It remains to be seen how this changes over development.

Development of Talker Representations in the Context of Language Development

Understanding children's use of talker information (and other paralinguistic cues) in language acquisition must be integrated with what we know about children's use of explicitly linguistic sound elements, such as phonemes. There are two obvious connections to be made with the broader speech-sound acquisition literature. On the one hand, one might expect children to improve more slowly at distinguishing voices than at identifying language-specific elements of meaning, as attention may be directed mostly to encoding meaning. On the other hand, gradual improvement in voice identification is consistent with recent work on children's gradual acquisition of multiple types of sound categories.

The possibility of poorer voice memory in preschool children than in adults raises the question of how voice information is encoded. Earlier work by Mann et al. (1979) and the current work suggest that children are not particularly good at discriminating novel talkers, though Mann et al.'s XAB task may have put strong demands on children's working memories. Evidence for poor voice memory conflicts with reports of acute sensitivity to speech acoustics (e.g., DeCasper & Fifer, 1980; Houston & Jusczyk, 2000; Kisilevsky et al., 2003; Schmale & Seidl, 2009). It is somewhat consistent with the hypothesis that developing speakers of a language progressively focus attention on lexically relevant acoustic information (Werker & Tees, 1984; see also Newman, 2008). Note that whether attention to phonemic characteristics suggests ignoring talker-related characteristics is an open question and may vary by task (see Werker & Curtin, 2005, for an account of attentional changes to speech over development).

On the other hand, the notion of gradually increasing voice identification has striking parallels in phonological development. Stager and Werker (1997; Swingley & Aslin, 2007) found that very young children (14 months) had difficulty learning novel words (bih, dih) distinguished by a speech-sound contrast that infants can discriminate at a much earlier age. One explanation for this effect in word learning is that it is difficult to map similar acoustic patterns onto external referents. The same may be true of talker-related acoustic patterns:

Though children can distinguish two somewhat similar acoustic patterns, it may be difficult to associate those patterns with specific individuals. This is certainly consistent with children's results in Experiment 3 (acoustically distinct talkers) compared to their results in Experiment 4 (less acoustically distinct talkers). Of course, adults in Experiment 4 exhibited the same pattern as children. It is also not clear whether both children and adults might use subtler talker differences when the voices and their owners are familiar. This would parallel work in phonological development as well, which shows that children can detect subtle (single-phoneme) differences in *familiar* words (Fennell & Werker, 2003; Swingley & Aslin, 2002) at the same age that learning subtly different novel words is difficult (Stager & Werker, 1997).

This perspective is consistent with an account by which listeners learn what elements of the speech signal they should direct attention toward in different contexts (e.g., Newman, 2008; Werker & Curtin, 2005). This includes exemplar-style theories of sound representation in language: that listeners store acoustic attributes of speech ''whole'' (Goldinger, 1998), including both phonemic variability and talker variability. What occurs over development, rather than a progressive narrowing of informational intake, is the emergence in the exemplar cloud of phoneme-related and talker-related regularities via perceptual learning. This is substantiated by parallel data on children's improvements through the single-digit years in recognizing accents (Floccia, Butler, Girard, & Goslin, 2009) and in understanding accented speech (Nathan, Wells, & Donlan, 1998). Only when these regularities can be accurately extracted can children map them to entities in the environment and direct attentional focus to different regularities. It may be that children are able to discriminate a large number of sound patterns but must go through a lengthy period of perceptual learning in order to decipher regularities in those acoustic patterns and map the regularities to referents, whether those referents are emotional states, accents, identities, or words.

Conclusion

Five experiments explored how preschool-aged children utilize talker-related acoustic characteristics in interpreting spoken language. Overall, children appear equivalent to adults in using talker information to predict outcomes consistent with their knowledge of a talker's color preferences. However, children may be less skilled than adults at linking voices to identities. Overall, the present data suggest that at least some paralinguistic cues—those indicating talker gender—can be used flexibly in sentence comprehension as early as 3 to 5 years of age.

References

Barr, D. J. (2008). Analyzing ''visual world'' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology*, *27*, 464–472.

Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, *107*, 87–99.

Bibyk, S., Ito, K., Wagner, L., & Speer, S. (2009). Children can use contrastive pitch accent in on-line processing. Paper presented at Boston University Conference on Language Development 34, Boston, MA.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707–729.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 687–696. doi:10.1037/0278-7393.30.3.687

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 521–533.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*, 207–239.

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 processing. *Cognitive Psychology*, *6*, 84–107.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience : The role of talker variation in lexical access. *Cognition*, *106*, 633–664.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*, 264–285. doi:10.1016/j.jml.2011.06.005

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(Pt. 2–3), 245–264.

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development*, *33*, 366–375.

Geiselman, R. E., & Crawley, J. M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning and Verbal Behavior*, *22*, 15–23.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 78–101). New York: Wiley.

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.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Perception*, *26*, 1570–1582.

Ito, K., Jincho, N., Yamane, N., Minai, U., & Mazuka, R. (2009). Use of emphatic pitch prominence for contrast resolution: An eye-tracking study with 6-year old and adult Japanese listeners. Paper presented at Boston University Conference on Language Development 34, Boston, MA.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 12577–12580.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within- and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*, 806–819.

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, *98*, 69–93.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*, 153–165.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*, 834–843.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech*, *50*, 125–142.

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, *25*, 343–365.

Newman, R. S. (2008). The level of detail in infants' word learning. *Current Directions in Psychological Science*, *17*, 229–232.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, *109*, 1181–1196.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*, *114*, 42–55.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, *109*, 2988.

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

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production. *Journal of the Acoustical Society of America*, *116*, 1905–1908.

Quam, C., & Swingley, D. (2012). Development in children's sensitivity to pitch cues to emotions. *Child Development*, *83*, 246–250.

Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, *111*, 372–377.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*, 608–635.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*, 441–452.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 915–930.

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*, 583–601.

Sekerina, I. A., & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*, *98*, 20–45.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*, 644–649.
Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*, 833–870.
Singh, L., White, K., & Morgan, J. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*, 157–178.
Singh, S., & Murry, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America*, *64*, 81–87. Retrieved July 20, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/712004
Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*, 574–608.
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.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*, 480–484.
Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*, 99–132.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.
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.
Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591.
Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters Part I: recognition of backward voices. *Journal of Phonetics*, *13*, 19–38.
Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters Part II: Recognition of rate-altered voices. *Journal of Phonetics*, *13*, 39–52.
Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 333–342.
Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*, 75–95.
Werker, J., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234. doi:10.1207/s15473341lld0102_4
Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

Appendix A

Error Rates in All Experiments

Experiment	Preferred-color errors %	Nonpreferred-color errors %	*t* value	*p* value
1	2.8	2.1	0.91	0.37
2	2.0	0.7	3.30	0.003**
3	1.6	0.8	1.35	0.19
4	1.4	1.1	0.55	0.59
5	2.4	2.5	0.12	0.91

Note. For nonpreferred-color errors, results were averaged across the two nonpreferred-color alternatives on a trial.
$^{**}p < .01$.

Appendix B

Correlations Between Participant Age and the Pink (Black) Looks on Female (Black-Preferring) Talker Trials Minus Pink (Black) Looks on Male (White-Preferring) Talker Trials.

Experiment	*n*	Pearson's *r*	*t* value	*p* value
1	22	−.220	1.01	.33
2 (1st person)	30	−.106	0.56	.58
2 (3rd person)	30	.186	1.00	.33
3	24	−.139	0.66	.52
4	30	−.003	0.02	.99
5	21	−.089	0.39	.70

Note. This score should be large and positive if the child is using talker information to execute looks to shapes. A small number of children (n = 3) are omitted due to missing age data.
$^{**}p < .01$.

Supporting information

Additional supporting information may be found in the online version of this article:

Appendix S1. Acoustic Measurements of Talkers.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.