Title

Preschoolers' flexible use of talker information during word learning

Author

Creel, Sarah C

Publication Date

2014-05-01

DOI

10.1016/j.jml.2014.03.001

Peer reviewed

Journal of Memory and Language 73 (2014) 81–98

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Preschoolers' flexible use of talker information during word learning

Sarah C. Creel *

UC San Diego, Department of Cognitive Science, 9500 Gilman Drive, Mail Code 0515, La Jolla, CA 92093-0515, United States

ARTICLE INFO

Article history:
Received 15 July 2013
revision received 1 March 2014
Available online 27 March 2014

Keywords:
Language development
Talker variability
Eye tracking
Sentence processing
Cue integration
Word learning

ABSTRACT

Previous research suggests that preschool-aged children use novel information about talkers' preferences (e.g. favorite colors) to guide on-line language processing. But can children encode information about talkers while simultaneously learning new words, and if so, how is talker information encoded? In five experiments, children learned pairs of early-overlapping words (geeb, geege); a particular talker spoke each word. Across experiments, children learned labels for novel referents, showing an advantage for original-voice repetitions of words which appeared to stem mainly from semantic *person-referent mappings* (who liked what referent). Specifically, children looked to voice-matched referents when a talker asked for their own favorite ("I want to see the geege") or when the liker was unspecified ("Point to the geege"), but they looked to voice-*mis*matched referents when a talker asked on behalf of the other talker ("Conor wants to see the geege"). Initial looks to voice-matched referents were flexibly corrected when later information became available (Anna saying "Find the geege for Conor"). Voice-matching looks vanished when talkers labeled the other talker's favorite referent during learning, possibly because children had learned two conflicting person-referent mappings: Anna-*likes*-geeb vs. Anna-*talks-about*-geege. Results imply that children's language input may be conditioned on talker context quite early in language learning.

Introduction

How do children encode multiple pieces of information during language learning? That is, how do children acquire the vast sets of associations between information sources—words, syntax, individual interlocutors—that allow adult-like language processing? Numerous studies have asked whether preschool- and young school-aged children can use multiple information sources in on-line language processing (e.g. Borovsky, Elman, & Fernald, 2012; Morton & Trehub, 2001; Trueswell, Sekerina, Hill, & Logrip, 1999). Fewer studies examine whether preschool-aged children can *learn* multiple cues to meaning simultaneously. Children, particularly in the preschool and early primary school age range, are notoriously weak at integrating or switching cues (color vs. shape, or syntax vs. referential content), especially when those cues conflict with each other (Morton & Trehub, 2001; Trueswell et al., 1999; Zelazo, Frye, & Rapus, 1996; see Morton & Munakata, 2002, for a neural network account). Does this fragile cue use imply that children will also have difficulty encoding multiple cues to linguistic meaning in language-learning situations? And does conflict between these newly-learned cues then cause confusion?

One case where preschool-aged children regularly experience multiple information sources is when they learn words from particular talkers. Do they encode talker information along with word information, and if so, how so? This question concerns not only cognitive flexibility,

* Fax: +1 858 534 1128.
E-mail address: creel@cogsci.ucsd.edu

http://dx.doi.org/10.1016/j.jml.2014.03.001
0749-596X/

but also the status of talker characteristics in the speech signal. Voice characteristics are typically regarded as "noise" in terms of recognizing speech sounds (though see, e.g., Johnson, Strand, & d'Imperio, 1999), yet voice characteristics are also "signal" in that they provide information about a speaker's individual or group identity. If the listener knows who is speaking, they can use their knowledge about that speaker's knowledge level (e.g. Koenig & Echols, 2003) and mental state (beliefs, desires, emotions) to enrich their understanding of what the speaker is talking about (see Akhtar, Carpenter, & Tomasello, 1996, for evidence that children as young as 2 years take into account what objects are novel to an adult interlocutor when learning novel words). The goal of the current study is to examine how and whether preschool-aged children learn novel word-referent and person-referent mappings from different talkers, and how children use such mappings when they conflict.

Development of talker processing

When considering how children might use talker information in language comprehension, a first point to address is whether listeners use talker characteristics during comprehension at all, or simply tune them out. It is important to distinguish between *acoustic/phonetic* talker characteristics and *semantic inferences* about talkers based on acoustic characteristics. One perspective is that talker variability is strained out of children's speech sound representations early on. While very young infants detect changes in both native and non-native speech sound contrasts, by 12 months (for many consonants, and earlier for vowels; e.g. Polka & Werker, 1994), children seem to ignore most non-native contrasts (e.g. Werker & Tees, 1984; though see, e.g., Narayan, Werker, & Beddor, 2010, for a case discrimination evident only after extensive native-language exposure [12 months]; and Ohde & Haley, 1997, for evidence of continued refinement of sound perception at 3–4 years). Some research is consistent with the viewpoint that talker variations are similarly ruled out. Kuhl (1979, 1983) successfully trained 6-month-olds to recognize vowel changes amidst changes in talker. Infants recognize word-forms over a change in talker by 10.5 months (Houston & Jusczyk, 2000), and over a change in accent around 12–13 months (Schmale, Cristià, Seidl, & Johnson, 2010; Schmale & Seidl, 2009). These results suggest that children learn very early to recognize words despite talker variability. Of course, recognizing a speech sound or a word despite variability does not necessarily entail *ignoring* talker (or accent) information. It may instead mean that children at 12 months (or well past that age) are still sensitive to talker information, but have learned to attend selectively to speech-relevant contrasts.

Evidence from older children (toddlers through preschoolers) supports the notion that sensitivity to talker variation is maintained well past infancy. Rost and McMurray (2009, 2010) showed that non-contrastive talker variability helped 14-month-olds distinguish highly-similar words in a word learning task. Richtsmeier, Gerken, Goffman, and Hogan (2009) found that preschoolers more accurately imitated nonsense words when they were heard in a variety of voices. In addition to these acoustically/phonetically-driven effects, preschool-aged and older children seem to draw semantic information from talker variation. Jerger, Martin, and Pirozzolo (1988) presented 3–6-year-olds with an auditory Stroop voice identification task, and found that children were slower to identify a talker as *mommy* or *daddy* when the talker and word were semantically incongruent (e.g., a female voice saying "daddy"). This interference pattern suggests that children process voices at a semantic level. Consistent with semantic processing of talker identity, 3- to 10-year-old children interpret nearly-identical sentences differently depending on who is speaking (Borovsky & Creel, in press; Creel, 2012), and similarly, they can make social decisions based on accent characteristics (Kinzler & DeJesus, 2013; Kinzler, Dupoux, & Spelke, 2007; see also Hirschfeld & Gelman, 1997).

By adulthood, listeners seem sensitive both to the acoustic properties and semantic implications of talker variability. While adults easily recognize words across voice changes, they are still sensitive to acoustic talker variation. Listeners are better at picking out a particular target in a series of speech sounds or words when the series is spoken by the same talker rather than by varying talkers (Magnuson & Nusbaum, 2007; Mullennix, Pisoni, & Martin, 1989; Nusbaum & Morin, 1992). Adults are better at learning second-language speech sounds when sounds are presented from multiple talkers or in multiple phonological contexts (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991). Listeners detect previously-presented words better when those words are repeated by the same speaker who originally spoke them (Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993). Adults can use talker information to distinguish otherwise similar-sounding words (Creel, Aslin, & Tanenhaus, 2008; Creel & Tumlin, 2011). These results suggest that adults show residual sensitivity to talker-varying acoustic properties.

Adults also show effects attributable to semantic encoding of voices. They encode gender information with spoken sentences (Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). They also show a larger semantic mismatch evoked potential when the voice is incongruous with sentence content (e.g., "I like to drink wine" in a child's voice; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). Knowledge of talker identity may even feed back to speech processing: the talker's apparent gender (Johnson, Strand, & d'Imperio, 1999) or dialect (Niedzielski, 1999) biases perception of what speech sound is being heard. Thus, adults process talker information at both acoustic and semantic levels.

Of course, it often is difficult to determine the level at which talker effects are taking place. One might break this down into *talker-specific word encoding, voice-referent mappings*, and *person-referent mappings* (Fig. 1). Note that the first two alternatives are essentially acoustic in nature, while the third (person-referent mapping) is a higher-level, more semantic use of talker information. Thus, when talker information affects adults' recognition and processing of the word *wine*, three things might be going on. First, the word *wine* may be represented *talker-specifically*, as having acoustic–phonetic characteristics consistent with adult voices (e.g. a fundamental frequency [f0] below

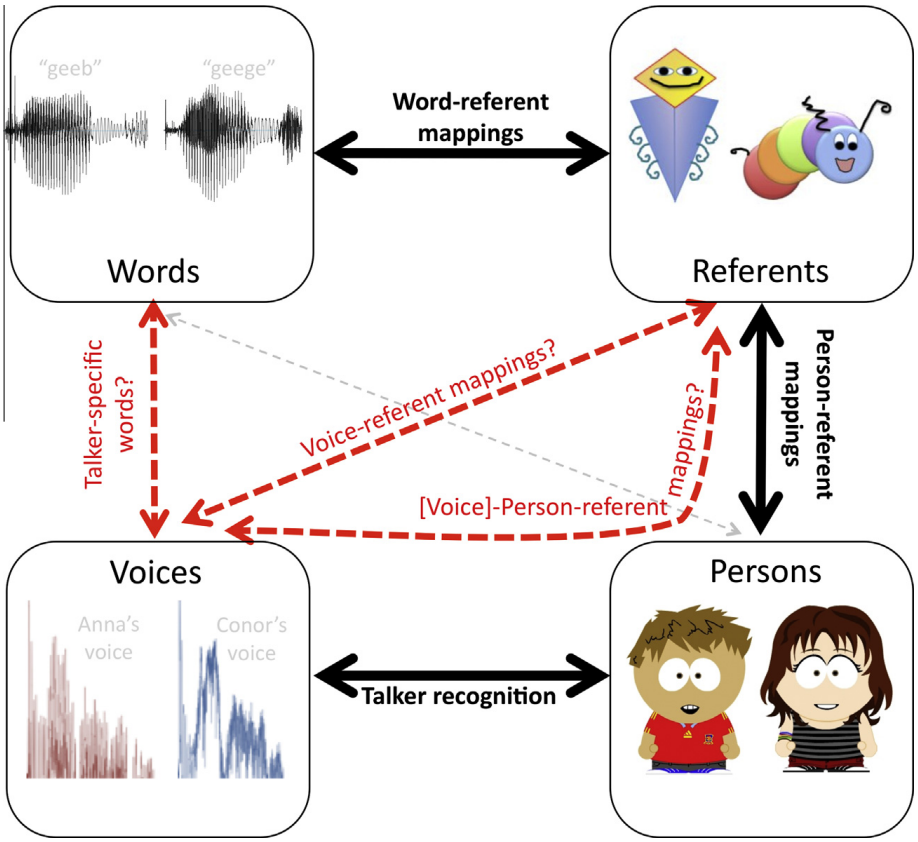

Fig. 1. Schematic of the types of mappings children might make when learning different words from different talkers. Thick arrows denote mappings with substantial empirical support. Dashed lines denote additional mappings explored in the current study. Light dashed line indicates a mapping that is not explored.

250 Hz). Second, the word *wine* may be associated with adult voice characteristics (*associated* with voices with f0 below 250 Hz). Third, the word *wine* may be *semantically* associated with adulthood—a person-referent mapping; in this case, voice information is just one way to activate semantic information about the speaker.

One might think that semantic use of talker information would be more deliberate or conscious, while acoustic uses of talker information would be more implicit. Interestingly, in studies where adults learned similar-sounding words either with or without a talker cue, they did not appear to use talker to great strategic advantage (i.e., they do not seem to deliberately exploit voice as an easy "shortcut" to the correct response): Creel et al. (2008) and Creel and Tumlin (2011) found little to no accuracy advantage for word pairs distinguished by who the talker was (e.g. "boog" was always spoken in a female voice and "booge" in a male voice), even though visual fixations were more rapid for talker-distinguished pairs relative to single-talker pairs (both "boog" and "booge" spoken by the same male voice). Creel and Tumlin also found evidence consistent with person-referent mappings: in some cases, participants looked to the correct referent before hearing the word at all. However, this pattern appeared only when the voice perfectly predicted the target, and explicit recall of those mappings was poor. This suggests that these were either weak person-referent mappings, or perhaps voice-referent mappings (associating voice acoustic characteristics with referents, rather than associating talker semantic characteristics with referents). However, it is impossible based on Creel and Tumlin's data to dissociate the two influences. In the current work, particular manipulations attempt to isolate different types of talker information.

Preschool children's encoding of talker information

The research reviewed above suggests that infants are overly-sensitive to "noise" talker characteristics (Houston & Jusczyk, 2000; Rost & McMurray, 2009, 2010), while adults are adept at using semantic talker information (e.g. Van Berkum et al., 2008) with residual sensitivity to talker characteristics (e.g. Palmeri et al., 1993). How does talker-semantic sensitivity emerge over development? It is known that preschoolers can access talker-related semantic properties flexibly in at least some contexts. Creel (2012) taught preschool-aged children that one character, Billy, liked blue (or white) things, while the other character, Anna, liked pink (or black) things. Children then saw sets of four shapes, and one talker requested a particular shape ("Can you help me find the star?"). Children looked to the shapes in that talker's favorite color, prior to target word onset ("star"), suggesting they were activating knowledge about the talker. This did not appear to result solely from voice-referent mappings (thinking "pink" whenever Anna spoke): when talkers sometimes asked for a shape *for the other character* ("Billy wants to see the star!" in Anna's voice), children looked to the other character's favorite color, not the talker's favorite color. This im-

plies children can switch between using voice information and third-person reference to determine whose color preferences are relevant.

In a related study, Borovsky and Creel (in press) showed that 3–10-year-olds and adults use voice information to access long-term knowledge about roles (e.g. pirate vs. princess) during sentence comprehension, integrating role knowledge with verb constraints. For instance, when a "princess" said "I want to hold the wand," children looked more toward the wand than any other picture before hearing the word "wand"—even though princess-associated (carriage) and holdable (sword) distractors were visible. This suggests that children use voice information to infer identity, and they then integrate this identity information seamlessly into sentence comprehension.

These two studies suggest that children aged 3–5 years (and older) are sensitive to talkers' identities. More specifically, they can use preexisting knowledge about individuals (things they like to hold) and integrate this knowledge with sentence constituents such as verbs (Borovsky & Creel, in press). They can also learn new information about individuals and use that in sentence comprehension (Creel, 2012)—that is, they can learn person-referent mappings. However, children regularly experience more complex sets of mappings, as depicted in Fig. 1, and it is not clear how (and in what circumstances) talker information is incorporated. Given children's limited cognitive resources, what set or subset of these mappings do they learn?

The current study

The current study places preschool-aged (3–5-year-old) children in learning situations which contain both novel word-referent and novel person-referent mappings. Previous studies have shown only that children can learn novel person-referent mappings and access those via voice information (favorite colors; Creel, 2012), or that children can use voice information to access known person-referent mappings (princesses like to hold wands; Borovsky & Creel, in press). No studies have considered whether children can *simultaneously* acquire word-referent and person-referent mappings that are accessible from voice information. Additionally, few studies have examined *how* children encode and use voice information to do anything other than recognize voices.

The study addresses three questions. First, do children use talker information to recognize words, and if so, how do they do so? Second, do children encode *and use* person-referent mappings even when learning new word-referent mappings? Third, what happens when voice cues (talker-specific words or voice-referent mappings), semantic cues (person-referent mappings), and phonological (word) cues are at odds with each other—how do children resolve the conflict?

In each of five experiments, children learned to recognize pairs of cartoon creatures (designated "referents" throughout) which had similar-sounding names. Test trials presented both cartoon referents side-by-side, and one referent was named by one of the training talkers. Children's accuracy (pointing responses) and eye movements to 2pictures were recorded.

Experiment 1 tested whether similar-sounding names that differed in the voice that spoke them (talker-distinguished pairs; male "mard" (/mɑɹd/), female "marv" (/mɑɹv/)) were recognized more quickly and accurately than names without a talker distinction (single-talker pairs; male "geeb" (/gib/), male "geege" (/gidʒ/)). Experiments 2–5 examined how children had encoded the talker information—acoustically, semantically, or both. Children again learned similar-sounding names, but all were talker-distinguished pairs, and talkers expressed preferences for different referents (for instance, "That's a geeb! I like the geeb best!"). Some test trials presented the original talker requesting a referent for herself/himself (Conor saying "I want to see the geeb!"), while on other trials a talker requested the other talker's "favorite" referent for that talker (Anna saying "Conor wants to see the geeb!"). Across experiments, two factors were manipulated. First was the timing of "liker cues:" first-person pronouns (I, me) or third-person nouns (Anna, Conor) indicating who wanted to see a particular referent. The second factor was the initial labeler's level of interest in the referent: in Experiments 2–3, the talkers labeled their own favorite creatures during learning trials, while in Experiments 4–5, each talker labeled the *other* talker's favorite creature during learning trials. This attempted to dissociate talker-specific words and voice-referent mappings (this creature is always heard with these voice acoustics) from person-referent mappings (Anna likes this creature).

Experiment 1

Experiment 1 asked whether preschool-aged children use talker information when recognizing novel words. The experiment was a simplified replication of earlier studies in adults (Creel & Tumlin, 2011; Creel et al., 2008). Children learned two pairs of onset-overlapping words: a single-talker pair, and a talker-distinguished pair. If children, like adults in previous studies, are sensitive to talker content in newly-learned words, then they should look more rapidly to targets in the talker-distinguished condition than in the single-talker condition. If this is due to talker-specific word representations, looks in the talker-distinguished condition should exceed looks in the single-talker condition after word onset. If this is due to voice-referent mappings, then looks in the talker-distinguished condition should exceed single-talker looks even earlier (before word onset).

Method

Participants

$N = 32$ monolingual English speaking preschool-aged children ($M = 4.17$ years, $SD = 0.60$; 19 female) took part.

Stimuli

Since one of the questions was whether children would encode words themselves talker-specifically, novel, phonologically-similar words were used: *geeb*, *geege*, *marv*, and *mard*. The logic of using novel words (as in Creel et al.,

2008; Creel & Tumlin, 2011) was that children, like adults, have heard familiar words in a variety of voices, while they will not have heard novel words in *any* voices, allowing tight control over talker-specific exposure. The logic of using words with a temporary phonological ambiguity was to delay the onset of phonological information that distinguishes the words. That is, if children are *not* storing talker-specific word representations, then visual fixations to the correct word cannot be generated until the end of the word. However, if children *are* storing talker-specific word representations, then visual fixations to the correct referent should be generated much faster on talker-distinguished trials (because talker information disambiguates these words earlier) than on single-talker trials. (Of course, as will become evident later, earlier looks on talker-distinguished trials can be interpreted in multiple ways.) As in Creel and Tumlin (2011), novel words contained all voiced segments, as voiced segments carry strong cues to talker identity (particularly f0 and vowel formant frequencies). The voiced segments in coda position also lengthened the words' vowels (relative to placing unvoiced segments in coda position), which lengthened the duration of word ambiguity.

One male talker ("Conor") and one female talker ("Anna") recorded four learning sentences, and an additional four testing sentences containing target words (top of Table 1). Word onset was 538 ms (SD = 190) after sentence onset. The average word point of disambiguation (POD) was at 423 ms (SD = 39) after word onset. Talkers' utterances were allowed to vary naturally, resulting in strong differences in mean fundamental frequency (f0) and f0 variability, and, in some experiments, differences in duration (Appendix A). As is typical of female vs. male voices, formant frequencies differed as well (Appendix A, Fig. A1). Numerous other factors may influence perception of voice quality, including individual listener differences (see, e.g., Kreiman, Gerratt, Precoda, & Berke, 1992).

Visual stimuli (examples in Fig. 2), used as referents for the novel words, were four colorful, distinct cartoon creatures created in Microsoft PowerPoint and exported as .jpgs. Creel and Jiménez (2012) and Creel (2014) have used these stimuli to study voice learning and word learning in this age group.

Design

Each child completed two learning-test sequences, once with a single-talker word pair and once with a talker-distinguished word pair (Table 2). Order of pair types (single-talker pair, talker-distinguished pair) was counterbalanced across children. Across children, each word pair (mard/marv, geeb/geege) was heard equally often as a single-talker or a talker-distinguished pair, and each word was heard equally often from the female talker and the male talker. Within a learning-test sequence, there were 16 learning trials (8 per word) and 16 test trials. On each learning trial (Fig. 2a), a cartoon creature moved onto the computer screen, paused in the center, and one talker labeled it. Each labeling event presented the creature's name twice. After being labeled, the creature moved back off the screen. Then the next learning trial occurred.

Learning trials occurred in a different random order for each participant, intermixing the different training phrases and the words labeled. After the first 8 trials in a learning block, a brief distractor sequence was presented (pictures of animals accompanied by entertaining sounds) to

Table 1
Learning and testing sentences.

Learning sentences	Testing sentences
Experiment 1	
Look at the XXXX! Do you see the XXXX?	Where's the XXXX?
A XXXX! That's a XXXX!	Point to the XXXX!
See the XXXX over there? Isn't it a nice XXXX?	Can you show me the XXXX?
It's a XXXX! Look at that XXXX go!	Find the XXXX!
Experiment 2	
The XXXX—yay, the XXXX is my favorite one!	[I/Char.] want[s] to see the XXXX!
That's a XXXX! I like the XXXX best!	Can you help [me/Char.] find the XXXX?
See that XXXX? I love the XXXX!	[I/Char.] [have/has] to see the XXXX! Where is it?
That XXXX is soooo cool! Go, XXXX!	Show [me/Char.] the XXXX!
Experiment 3	
The XXXX—yay, the XXXX is my favorite one!	Point to the XXXX for [me/Char.]!
That's a XXXX! I like the XXXX best!	Can you find the XXXX for [me/Char.]?
See that XXXX? I love the XXXX!	Where's the XXXX? [I/Char.] [have/has] to see it!
That XXXX is soooo cool! Go, XXXX!	Show the XXXX to [me/Char.]!
Experiment 4	
[Char.] likes the XXXX—the XXXX is her/his favorite one!	Point to the XXXX for [me/Char.]!
[Char.] likes the XXXX best! That's a XXXX!	Can you find the XXXX for [me/Char.]?
Does [Char.] see that XXXX? She/He loves the XXXX!	Where's the XXXX? [I/Char.] [have/has] to see it!
[Char.] thinks the XXXX is so cool! Look at that XXXX go!	Show the XXXX to [me/Char.]!
Experiment 5	
[Char.] likes the XXXX—the XXXX is her/his favorite one!	[I/Char.] want[s] to see the XXXX!
[Char.] likes the XXXX best! That's a XXXX!	Can you help [me/Char.] find the XXXX?
Does [Char.] see that XXXX? She/He loves the XXXX!	[I/Char.] [have/has] to see the XXXX! Where is it?
[Char.] thinks the XXXX is so cool! Look at that XXXX go!	Show [me/Char.] the XXXX!

Note: XXXX = word; Char. = character's name, either Anna or Conor depending on the trial.

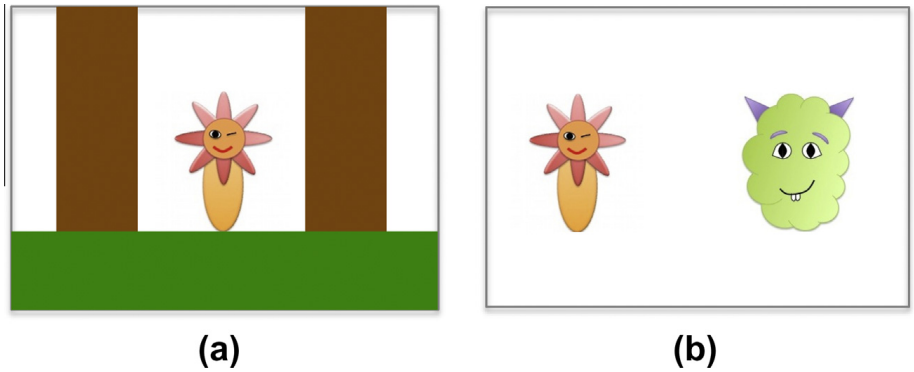

Fig. 2. Example visual displays from (a) learning trials and (b) test trials.

Table 2
Trial structures across experiments.

Trial *n*	50% of ppts.	Other 50%
Experiment 1		
16 learning	single-talker	talker-distinguished
16 testing	single-talker	talker-distinguished
16 learning	talker-distinguished	single-talker
16 testing	talker-distinguished	single-talker

Trial *n*	Trial types
Experiments 2–3	
16 learning	talker-distinguished labeling
16 testing	50% first-person, 50% third-person
16 learning	talker-distinguished labeling
16 testing	50% first-person, 50% third-person
Experiments 4–5	
16 learning	third-person talker-distinguished labeling
16 testing	50% first-person, 50% third-person
16 learning	third-person talker-distinguished labeling
16 testing	50% first-person, 50% third-person

maintain child interest. After 8 more learning trials, 16 test trials (Fig. 2b) occurred in a different random order for each participant. Each test trial presented the two creatures stationary, side by side on the screen (side was counterbalanced within a test block). The original talker who had labeled the creature continued to label it during test trials. Thus, on talker-distinguished blocks, talker predicted the correct answer, while on single-talker blocks, talker did not predict the answer because the same talker spoke both words. Another distractor sequence played after the first full learning-test sequence. Then the second learning-test sequence began. During test phases, children's accuracy (points to pictures) was recorded, along with visual fixations to pictures. Pointing data were recorded via mouse click by an experimenter sitting next to the child. When points were ambiguous, children were prompted to clarify their points. On rare occasions, when children continued to point ambiguously or refused to point, the experimenter queried *both* pictures ("Is it this one? [Point left] Is it this one?" [Point right]) and recorded a response only if the child verified one but rejected the other.

Equipment

Fixations were recorded by an Eyelink 1000 Remote eye tracker (SR Research, Mississauga, ON), which sampled gaze position every 4 ms. The eye tracker was calibrated just prior to the experiment using standard 5- or 9-point calibration routines. The experiment was run using the PsychToolBox 3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002) for Matlab. Children sat in an unbuckled car seat to help maintain a consistent distance from the eye tracker.

Results

Accuracy

Children achieved reasonable accuracy (Fig. 3, far left) on both talker-distinguished ($M = .721$, $SD = .231$; above chance, $t(31) = 5.06$, $p < .0001$) and single-talker trials ($M = .664$, $SD = .191$; $t(31) = 3.84$, $p = .0006$). Throughout, accuracy was empirical-logit transformed prior to analysis to correct for non-normal distribution. To assess whether talker as a distinguishing factor affected accuracy, an analysis of variance (ANOVA) was computed on transformed accuracy with Trial Type (talker-distinguished, single-talker) as a within-participants factor and Block Order (talker-distinguished block first, single-talker block first) as a between-participants factor. Trial Type did not approach significance, suggesting that accuracy did not differ between talker-distinguished and single-talker trials. The lack of an accuracy effect implies that children were not using talker identity deliberately to strategic advantage. Block Order and the Block Order × Trial Type interaction did not approach significance.

An alternative account of performance on talker-distinguished trials is that children are not learning the

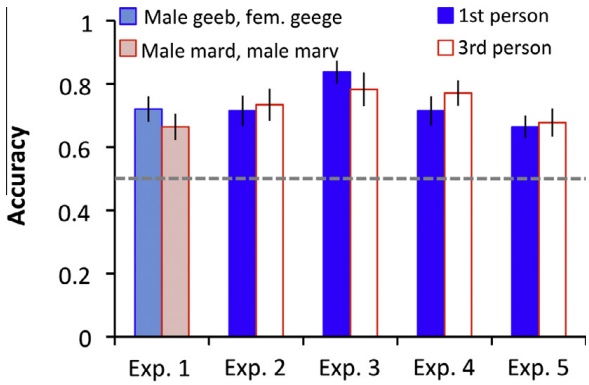

Fig. 3. Accuracy with standard errors. All bars exceed chance (dotted line), $p \leqslant .0006$.

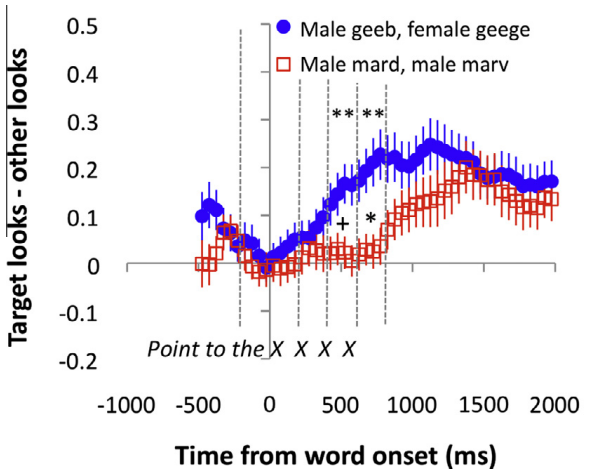

Fig. 4. Experiment 1, visual fixations from word onset with standard errors. A sample sentence indicates that 0 = word onset (note that throughout, the preceding word(s) and the exact target word duration are not aligned). XXXX = target word. Asterisks above solid circles compare circles to chance; asterisks between lines compare solid circles to hollow squares. [+]Bonferroni marginal, [*]Bonferroni significant $p < .05$, [**]Bonferroni significant, $p < .01$.

word-forms at all, but are learning a much-easier voice-creature mapping. That is, they are "cheating" at word learning by associating the voice with the creature rather than associating the creature with the word form itself. However, accuracy in talker-distinguished trials was only 72%. This is much lower than children's 92% accuracy on a *talker*-learning task with male vs. female voices (Creel & Jimenez, 2012). In Creel and Jiménez's task, preschool-aged children simply had to remember each creature's voice (one male voice, one female voice), rather than remembering a particular word for each creature. A *t*-test assuming unequal variances indicated lower performance here for talker-distinguished word pairs ($t(32.65) = 3.96$, $p = .0004$) than for talker learning in Creel and Jiménez. This suggests that children were not, by and large, "cheating" in the relatively-challenging task of learning two phonologically-similar labels by using talker information (arguably a more prominent acoustic difference) instead.

Visual fixations from word onset

An inspection of Fig. 4 suggests that talker-distinguished words were recognized more rapidly than single-talker words. Throughout, looks to target and competitor underwent e-logit transformation prior to analysis to correct for non-normal distribution. Trials with both correct and incorrect responses were included.[1] Next, a target advantage score—target looks minus competitor looks—was computed to measure children's relative looking preference for the target (positive target advantage), competitor (negative target advantage), or neither (target advantage near zero). Looks were analyzed in a time window spanning 1000 milliseconds (ms), beginning 200 ms before word onset and ending 800 ms post-word onset. This time window encompassed a brief period prior to word onset, as well as time intervals that might contain visual fixations based on auditory cues (typically assumed to begin no earlier than 200 ms after the relevant external cue; see Hallett, 1986), including the end of the word: average final consonant onset was 423 ms ($SD = 39$ ms), such that final-consonant-driven looks should begin appearing around 623 ms, early in the last time window. Much later time windows are a bit problematic to investigate because children's pointing responses tend to obscure the eye tracker camera, which artificially deflates looking proportions. Throughout, all children in the sample looked to one of the two pictures 50% or more of the time during the analyzed time window.

An analysis of covariance—an ANOVA with a continuous factor—was computed on target advantage scores with Trial Type and Block as discrete within-participants factors, and Time Window (−200 to 0 ms, 0–200 ms, 200–400 ms, 400–600 ms, 600–800 ms) as a mean-centered continuous factor. An effect of Time Window ($F(1,30) = 8.65$, $p = .006$) indicated an increase in target advantage over the analyzed time period. An effect of Trial Type ($F(1,30) = 4.33$, $p < .05$) resulted from greater target advantage in talker-distinguished trials than single-talker trials. Finally, a Trial Type × Time Window interaction ($F(1,30) = 4.66$, $p = .04$), indicated a more rapid increase in looks in talker-distinguished trials than in single-talker trials. Individual *t*-tests were computed comparing the two trial types in each time window (Bonferroni-corrected for five tests; see Fig. 4), indicating divergence of the two trial types by the 600–800 ms time window. Individual *t*-tests also compared each trial type to chance in each window (again, Bonferroni-corrected for five tests; see Fig. 4), indicating that only the talker-distinguished trials began to exceed chance during the analyzed time window, around 400 ms. This is consistent with children using talker-specific word representations.

Discussion

Children learned talker-distinguished word pairs and single-talker word pairs with comparable accuracy, but showed a looking-time advantage for talker-distinguished word pairs. This mirrors experiments with adults (Creel & Tumlin, 2011; Creel et al., 2008), where learners were faster to fixate to referents with labels distinguished by talker information. Also like adults (Creel & Tumlin, Experiment 2), visual fixations to talker-distinguished targets appeared after word onset. This is consistent with encoding talker-specific word representations. However, this could also reflect learning of person-referent or voice-referent mappings, rather than talker-specific-word representations. Regardless, data are consistent with encoding of talker information during word learning. The following experiments attempted to distinguish whether children were storing talker-specific words or voice-referent mappings—low-level perceptual explanations of talker-specificity effects—vs. encoding person-referent mappings (who likes what), a higher-level semantic explanation.

[1] This was done for two reasons. The first reason is comparability with studies of even-younger children than tested here; in those studies, error trials are unknown (because no overt decision measure can be obtained). The second reason is that discarding error trials can give the false appearance of earlier recognition. This may occur because, on some proportion of both correct and error trials, children are guessing, and may be biased to select whatever they are looking at first. Removing only the guessing trials where children were in error may thus inflate apparent early looks to the correct referent.

Experiment 2

When do children form person-referent mappings? In this experiment, children were introduced to two characters, Anna and Conor, and were explicitly told that each talker preferred a particular cartoon referent. For instance, Conor might say "A geege! Yay, the geege is my favorite one!" Children were tested with first-person trials ("I want to see the…") which contained congruent voice information, as well as third-person trials ("Anna wants to see the…") which contained incongruent voice information.

If children are learning *talker-specific words*, then they should look to the correct referent fairly soon after target word onset. If they are instead learning *voice-referent mappings*, then they should look to the correct referent rapidly—*before* word onset—on first-person trials (where the voice matches the one at learning), but should look toward the incorrect referent on third-person trials (where the voice *mismatches* the one at learning). Finally, if children are learning *person-referent mappings* (talkers' preferences) in addition to word-referent mappings, they should look toward, and select, the correct referent on first-person test trials based on voice cues to who the liker is, and third-person test trials based on proper name cues ("Anna") to the liker.

Method

Participants

$N = 32$ new monolingual English speaking preschool-aged children ($M = 4.33$ years, $SD = 0.59$, 17 female) took part.

Stimuli

Talker pictures were generated using the South Park Studio web site (http://www.sp-studio.de/). Talkers were the same as Experiment 1. However, learning and testing sentences were different. In each learning sentence, the talker expressed strong personal preference for the named referent (Table 1). This was designed to encourage children to learn person-referent mappings, in addition to learning talker features of the word itself. Word onset was 940 ms ($SD = 256$) after sentence onset. The average word POD was at 414 ms ($SD = 44$) after word onset.

Design

This was similar to Experiment 1, except that both learning-testing iterations contained talker-distinguished pairs (Table 2). Across children, each word pair (mard/marv, geeb/geege) was heard equally often in the first or second learning-testing iteration, and each word was heard equally often from the female talker and the male talker. Also, to emphasize the relevance of talkers' preferences, two introduction trials preceded each set of learning trials and test trials. On each introduction trial, Anna (or Conor) appeared in cartoon form (see Fig. 1) in the center of the screen, and asked the child to assist them in selecting their "favorite animal." Importantly, this was the only time that children actually saw the cartoon talkers. Thus, on both learning and test trials, the only indication of the talker's identity was their voice. Test trials in both blocks were evenly split between original-talker (first-person) trials, and other-talker (third-person) trials. Note that the liker was always the correct liker for the referent requested, even though the voice only matched training half of the time.

Results

Accuracy

As in Experiment 1, accuracy was good, but not at ceiling (Fig. 3, middle left) on both first-person (original-talker) trials ($M = .715$, $SD = .277$; greater than chance, $t(31) = 4.44$, $p = .0001$) and third-person (other-talker) trials ($M = .734$, $SD = .289$; $t(31) = 4.40$, $p = .0001$). An ANOVA assessed accuracy, with Trial Type (first person, third person) and Block (first, second) as within-participants factors. A main effect of Block ($F(1,31) = 4.27$, $p < .05$), suggested higher accuracy in the second block ($M = .754$, $SD = .249$; vs. $M = .695$, $SD = .310$, in the first block). However, there was no effect of Trial Type nor a Trial Type × Block interaction. That is, children showed no decrement in accuracy when the target word was presented in the "wrong" voice. This suggests that phonological information (the ends of the target words) and/or the "liker" (pro)noun (I, me, Anna, Conor) were more important to children than talker-specific word information or voice-referent mappings.

Visual fixations from word onset

Eye tracking results (Fig. 5) suggested that children used liker-identifying information to guide visual fixations to the target prior to word onset. An ANOVA was computed on target advantage scores with Trial Type (first person, third person), Block, and Time Window as factors. There was an effect of Time Window ($F(1,31) = 8.01$, $p = .008$), indicating an increase in target advantage over the analyzed time period. No other effects or interactions

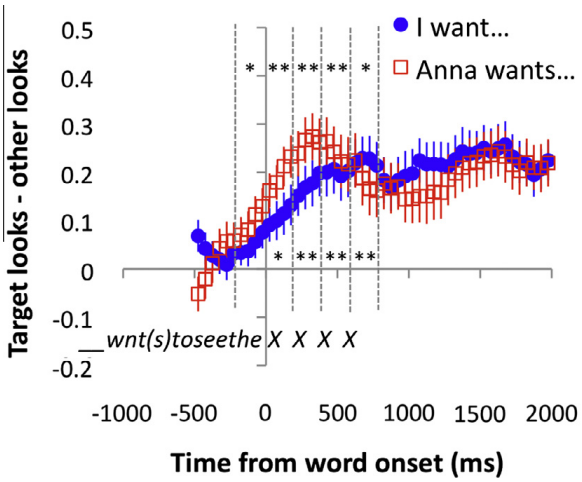

Fig. 5. Experiment 2, looks from word onset. Throughout, ___ refers to the first person pronoun (I/me) or the third-person proper name (Anna/Conor). Asterisks above hollow squares compare squares to chance; asterisks below filled circles compare solid circles to chance. All significant comparisons involved each trial type vs. chance (that is, 0 target advantage); the two trial types did not differ from each other. *Bonferroni ⩽ .05, **Bonferroni ⩽ .01.

approached significance. Individual *t*-tests also compared each trial type to chance in each window (Bonferroni-corrected for five tests; see Fig. 5), indicating that both trial types exceeded chance looks in the 0–200 ms time window (earlier, for third-person trials)—that is, earlier than could be predicted based on information in the word itself, and earlier relative to word onset than in Experiment 1's talker-distinguished condition. One possibility is that there might be a "dip" in looks after word onset in third-person trials, due to a mismatch to a talker-specific word representation. However, no such dip appeared, providing no evidence of talker-specific effects limited to the words themselves.

Visual fixations from sentence onset

While there do not seem to be talker-specific word recognition effects, it is of interest to determine what exact pre-word information children are using to identify the liker. On the one hand, they might be using early voice cues (indicating voice-referent mappings). On the other hand, they might be using liker cues ("I" or "Anna"; indicating person-referent mappings). Therefore, looks were realigned to sentence onset (Fig. 6). Note that for two test sentences, liker-cue onset (noun or pronoun) is at the beginning of the sentence, while for the other two sentences, the liker cue begins later. Accordingly, Liker Onset (initial, non-initial) was included as a factor in the analysis. Formally, an ANOVA on target advantage was calculated with Time Window (200–400, 400–600, 600–800, 800–1000, 1000–1200), Trial Type, Liker Onset, and Block as within-participants factors. An effect of Time Window ($F(1,31) = 14.92$, $p = .0005$) indicated an increase in target advantage over time. An effect of Liker Onset ($F(1,31) = 9.67$, $p = .004$) resulted from greater target advantage when the liker cue occurred sentence-initially, indicating use of liker cues. Finally, an effect of Trial Type ($F(1,31) = 4.29$, $p < .05$) suggested higher target advantage for first-person than third-person trials. This might be due to voice-referent mapping, as discussed below. No other effects reached significance.

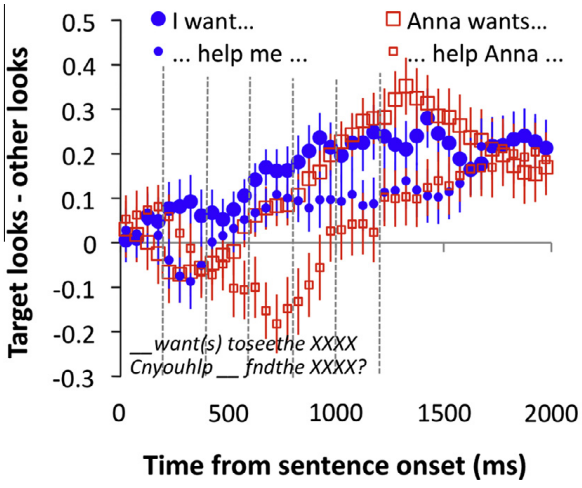

Fig. 6. Experiment 2, looks to pictures aligned with sentence onset. Large points indicate trials with sentence-initial liker cues, and small points indicate trials with non-initial liker cues.

Discussion

As in Experiment 1, children learned words accurately. Unlike Experiment 1, looks to the target suggested that children had stored not only word-referent mappings, but also person-referent mappings. Specifically, target looks surpassed looks to the competitor as early as 0–200 ms after word onset, which is before phonological information in the word itself could guide looking patterns. Further, these early looks were observed both for first-person (original-talker) trials and third-person (other-talker) trials, and appeared sooner after sentence onset for liker (pro)noun-initial sentences. These findings suggest that children had learned person-referent mappings. Interestingly, there was no evidence of conflict between liker cues and voice cues in the region of the target word, as might be expected on an account of talker-specific acoustic encoding of words (e.g. Creel & Tumlin, 2011; Creel et al., 2008). If so, there should have been a deflection in looks to the voice-matching picture after word onset, and this did not occur.

These results suggest that children may predominantly encode talker information at a semantic level, that is, they form person-referent mappings. Does this mean that children do not form direct voice-referent mappings at all, or that they do not store talker-specific word representations? Not necessarily. The liker cues that children used to show person-referent mapping had one important advantage: order. These cues (I/me, Anna/Conor) always occurred prior to the target words. This may have given person-referent mappings an advantage over talker-specific word representations, and may have short-circuited any weaker voice-referent mappings. Consistent with this, a close look at Fig. 6 suggests that, when the liker cue was not sentence-initial, children may have been responding based on voice match, with correct looks on first-person trials (small filled circles), and initially *incorrect* but voice-matched looks on third-person trials (small open squares). The next experiment thus tested whether voice match might have a stronger influence—either in terms of talker-specific words or voice-referent mappings—if target words occurred prior to liker cues.

Experiment 3

This experiment replicated Experiment 2, but reversed the order of liker cues and words on test sentences so that talker-specific word information came first. If children flexibly use the cue that comes first, they should use talker-specific words or voice-referent mappings in the next experiment. That is, they should look to the correct referent on first-person trials, and to the incorrect referent on third-person trials. If they are learning talker-specific word representations, this looking pattern should not emerge until some time after word onset. If they are learning voice-referent mappings, then looks should emerge prior to word onset. However, if children are using talker information only in the service of person-referent mappings, they will not know whose preferences to invoke until after they have heard the word itself. Therefore, they should not look to the correct referent until the *end* of the word arrives—providing phonological cues to the correct

referent—or until even later, when liker cues arrive (after the word ends).

Method

Participants

$N = 32$ preschool-aged children ($M = 4.76$ years, $SD = 0.39$, 19 female) from the same population as before took part.

Stimuli

Visual stimuli were identical to previous experiments. Learning sentences were the same as those in Experiment 2. The same two speakers recorded new test sentences. Voice cues were correct (i.e., consistent with learning) on first-person trials but mismatched learning on third-person trials. Test sentences differed from Experiment 2 in that the *order* of word cues and liker cues was reversed (Table 1). As before, the liker in test sentences was always consistent with learning.

Word onset was 487 ms ($SD = 114$) after sentence onset. Average word POD was 382 ms ($SD = 67$) after word onset. Mean inter-onset interval for words and either first-person pronouns or names was 719 ms ($SD = 123$) on first-person trials, and 680 ms ($SD = 187$) on third-person trials, leaving time for voice-specific word effects to emerge prior to onset of person information.

Procedure

Aside from the change in test sentences, this matched Experiment 2.

Results

Accuracy

Children performed with moderate accuracy (Fig. 3, middle). Both first-person ($M = .838$, $SD = .207$, $t(31) = 8.64$, $p < .0001$) and third-person trials ($M = .783$, $SD = .305$, $t(31) = 4.95$, $p < .0001$) exceeded chance accuracy. In an ANOVA on transformed accuracy with Trial Type and Block as factors, no effects approached significance.

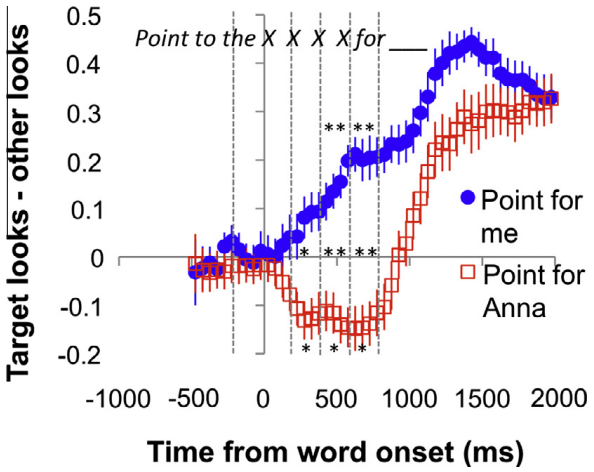

Fig. 7. Experiment 3, visual fixations from word onset. Asterisks above solid circles compare circles to chance; asterisks below hollow squares compare squares to chance; asterisks between lines compare solid circles to hollow squares. *Bonferroni-corrected $p < .05$, **Bonferroni-corrected $p < .01$.

Visual fixations from word onset

Results (Fig. 7) strikingly indicated that children looked to the talker's favorite picture on all trials, and then corrected on third-person (other-talker) trials when they reached the end of the word (see next analysis). An ANOVA was computed on target advantage scores with Trial Type, Block, and Time Window as factors.

An effect of Trial Type ($F(1,31) = 21.77$, $p < .0001$) indicated higher target advantage for first-person trials. Time Window did not approach significance, but the Trial Type × Time Window interaction was significant ($F(1,31) = 17.39$, $p = .0002$). Individual *t*-tests compared the two trial types in each time window (Bonferroni-corrected for five tests; see Fig. 7), indicating that the two trial types differed from each other by 200–400 ms. Individual *t*-tests also compared each trial type to chance in each window (Bonferroni-corrected for five tests; see Fig. 7), indicating that while first-person trials exceeded chance by 400–600 ms, target advantage on third-person trials was significantly *below* chance by 200–400 ms, with more looks to the voice-matched picture. The earliness of these voice-specific effects suggest that they were likely driven by voice-referent mapping rather than talker-specific word representations. However, consistent with the high accuracy on third-person trials, third-person looks rebounded in the positive direction by the end of the trial. A marginal effect of Block ($F(1,31) = 3.78$, $p = .06$) indicated higher target advantage overall in Block 1, but Block did not interact with any other factor, suggesting that the magnitude of the Trial Type effect did not change across blocks. No other effects approached significance.

Visual fixations from liker-cue (noun or pronoun) onset

It is clear that children are using voice information very early. However, it is less certain what drives the correction on third-person trials. Is it possible that children were not even learning the words for pictures, but were simply

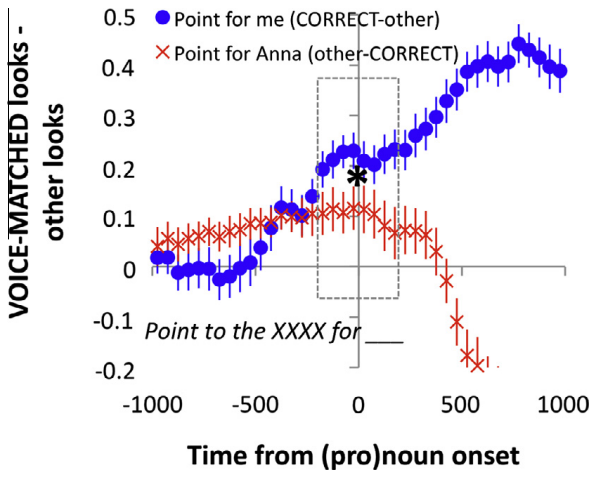

Fig. 8. Experiment 3, looks to the voice-matched picture minus looks to the voice-mismatched picture aligned at the onset of the liker noun or pronoun. Note that for the third-person (Point for Anna) trials, correct looks are negative. *$p < .05$.

learning who liked what? That is, were children looking to the correct referent on third-person trials based solely on the noun (Anna, Conor) or pronoun (I, me) cue to the liker, or the final sound of the word itself? It is important to establish whether children used word information because only then does it make sense to say that children are learning multiple information sources (both words and who likes what). To assess this, looking proportions were recentered to liker onset, and were rescored as looks to the voice-matched picture minus looks to the voice-mismatched (but, on third-person trials, the ultimately correct) picture. This amounts to inverting the third-person trial target advantage over the *x*-axis (Fig. 8).

As long as children are using only voice cues, third-person trial voice looks should perfectly track first-person trial voice looks (Fig. 8). However, when children begin to use other information (the end of the word, the liker [pro]noun, or both), third-person voice-matched looks should drop. The question was at what time point voice-based looks began to "correct" (i.e., when children began to look away from the voice-matched picture). If third-person voice looks start to differ before liker information is available, then this suggests that phonological information (the end of the word) is driving looks away from the incorrect referent. Target advantage for voice-matched pictures was assessed in the brief time window before liker-cue looks were possible: −200 to +200 ms after liker word onset, in an ANOVA with Trial Type and Block as factors. Sure enough, target advantage on third-person trials differed significantly at time windows earlier than was possible based on eye movements cued by the liker (pro)noun (that is, 200 ms after pronoun onset and later; $F(1,31) = 6.78$, $p = .01$). This suggests that children were beginning to correct their visual fixations on third-person trials based on word-final information, which in turn implies that they had actually learned the words themselves. It remains possible that liker (pro)noun information also contributed once it was available.

Discussion

While children in Experiment 3 maintained high accuracy, they showed dramatically different looking patterns relative to Experiment 2 when cues to the liker were delayed. Specifically, they initially looked more to the voice-matched picture, which was correct on first-person trials but incorrect on third-person trials. Voice-based looks appeared quite soon after word onset, making it possible but quite unlikely that these looks reflected talker-specific word representations. Rather, these results suggest that children encoded voice-referent mappings, which only became evident in the temporary absence of liker cues. Further, the magnitude of the visual fixation effect did not diminish from the first block of trials to the second block of trials. That is, children did not learn to ignore uninformative voice information in selecting the target.

It is noteworthy that children recovered readily after making erroneous initial assumptions on third-person trials. Accuracy was not impaired on these trials, and visual fixations rapidly veered toward the correct referent once the word was heard in full. This implies that not only can word-referent mappings and voice-referent mappings be learned concurrently, but that when these cues seem to conflict, the conflict is resolved correctly. This is a different pattern of cue conflict resolution than found by Trueswell et al. (1999) for 5-year-olds' use of pragmatic cues.

These results suggest that children do use voice-referent mappings to begin making hypotheses about what picture will be named, without waiting for liker cues. However, it is possible that this apparent use of voice-referent mappings actually represents an instance of *person-referent* mappings. That is, children may still be using voice cues to identify the person, and making a person-referent mapping, based on an assumption that the liker herself is making the request. This generates looks to the talker's preferred picture until children get word information or third-person cues ("Anna") to the contrary.

Why might children assume that talkers speak for themselves? Did they learn this contingency in the learning phase, where talkers did consistently speak for themselves, or do they have a preexisting bias to assume that talkers speak for themselves? This is not clear. The persistence of the voice-match effect into the second block of trials, well after the correlation of liker and voice was broken, suggests it may be a robust heuristic that children use to interpret spoken language. Nonetheless, if children are simply learning the speak-for-oneself pattern *during the experiment*, then they should be equally able to learn the opposite—speak-for-other—if each talker routinely labels the other talker's favorite during the learning phase.

Experiment 4

This experiment attempted to distinguish whether children's voice-driven looks in Experiment 3 reflected person-referent mapping or voice-referent mapping, by putting liker cues and voice cues in opposition during learning trials. Each talker labeled the other talker's favorite picture during learning. This meant that the potential voice-referent mapping was to the voice of the person who preferred the other referent, rather than to the liker. If children are using person-referent mappings with a stable assumption that talkers speak for themselves, then the liker's voice should generate looks to their preferred picture despite a voice mismatch to training. Like Experiment 3, this would generate above-chance looks on first-person trials and below-chance looks on third-person trials. However, if children are learning voice-referent mappings, or if they can flexibly learn that each talker speaks for the other, or both, then the *voice* associated with a picture should generate looks to that picture, even though it is not the liker's voice. These factors, either alone or in combination, would generate above-chance looks on third-person trials but below-chance looks on first-person trials, the inverse of the pattern from Experiment 3.

Method

Participants

$N = 32$ preschool-aged children ($M = 4.70$ years, $SD = 0.65$, 11 female) from the same pool as in previous experiments took part.

Stimuli

Visual stimuli matched Experiment 3. Test phrases also matched Experiment 3, with the target word occurring before the liker's name. However, learning phrases (Table 1, bottom) differed from Experiment 3 in that here, each talker labeled the *other character's* favorite creature.

Procedure

This matched Experiments 2–3. Note that talker-introduction trials remained unchanged, in that each talker asked the child to "help find my favorite animal," not to find the other talker's favorite animal.

Results

Accuracy

Accuracy was comparable to that in other experiments (Fig. 3, middle right), exceeding chance on both first-person test trials ($M = .715$, $SD = .264$, $t(31) = 4.54$, $p < .0001$) and third-person test trials ($M = .771$, $SD = .226$, $t(31) = 6.49$, $p < .0001$). An ANOVA on transformed accuracy assessed effects of Person and Block. No effects approached significance.

Visual fixations from word onset

Visual fixations did not diverge from chance until well after word onset (Fig. 9). In previous experiments, children used voice differences or third-person reference to visually fixate the target before words were differentiated. Here, though, *neither* trial type showed significant looks to the target until after word offset.

To assess this observation statistically, an ANOVA was computed on target advantage with Trial Type, Block, and Time Window as factors. The only effect approaching significance was Trial Type ($F(1,31) = 3.43$, $p = .07$), resulting from slightly-higher target advantage for first-person trials. Individual *t*-tests compared the two trial types in each time window, and compared each trial type to chance in each window (all Bonferroni-corrected for five tests; see Fig. 9). There was a hint of greater target advantage for first-person than third-person trials in the 200–400 ms time window, but no effects were significant after Bonferroni correction.

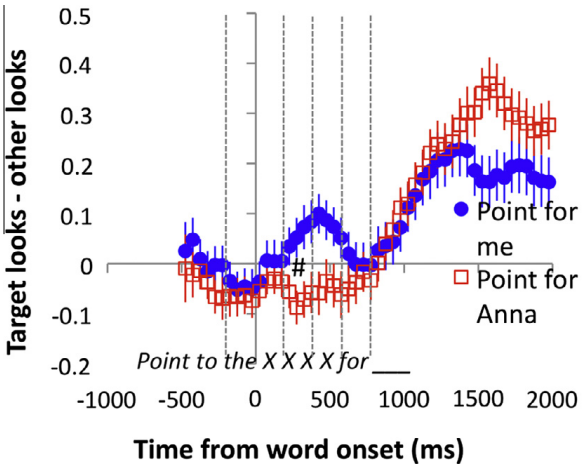

Fig. 9. Experiment 4 visual fixations. $^{\#}p = .03$, comparison between circles and squares, not significant after Bonferroni correction.

Discussion

Unlike previous experiments, children appeared to make relatively little use of voice-referent mappings *or* person-referent mappings. The marginal hint of greater looks to the liker-matched pictures on first-person trials would be consistent with person-referent mappings plus an assumption that likers will ask for their own favorites, but the strong voice-match effect observed in Experiment 3 has largely disappeared. Nor do results support flexible learning of a speak-for-other pattern, or voice-referent mapping: in both of those cases, third-person looks should have risen above chance early, while first-person looks should have dipped below chance.

Given the robust effects in Experiment 3, what do these results mean? One possible account of the results is that children are still storing person-referent mappings and assume that talkers will speak for themselves, but they also learn voice-referent mappings. In the current experiment, these two information sources were in opposition, and so canceled each other out. Another possibility is that two *semantic* associations are canceling each other out. That is, children have learned to associate each talker with both referents; Anna *likes* the **geeb**, but Anna *talks about* or *knows about* the **geege**. When they hear Anna's voice, they semantically activate both referents, generating equivalent looks to each in the absence of other information (the disambiguating phoneme in the word, or the liker noun or pronoun).

One last explanation of these results, if true, renders both of the above possibilities moot. Specifically, perhaps children simply failed to learn person-referent mappings—who likes what—due to the unusual, and possibly pragmatically awkward, learning situation. That is, perhaps children were confused by talkers labeling each other's favorite creature rather than their own favorite creature, and this confusion blocked learning of person-referent mappings and voice-referent mappings. This possibility was assessed in the final experiment.

Experiment 5

The purpose of this final experiment was to assess whether the data pattern in Experiment 4—apparent failure to use talker or liker cues—resulted from failure to learn person-referent mappings. Children heard learning trials as in Experiment 4, but at test, "liker" information was earlier in the sentence, as in Experiment 2. If children cannot form person-referent mappings when instructed in third-person form ("Conor likes the marv"), then early liker cues should not aid them in selecting the target. That is, the visual fixations should be identical to Experiment 4. However, if children *can* learn person-referent mappings from third-person descriptions, then visual fixation patterns should be similar to Experiment 2, where children fixated

the correct referents before phonological cues (the end of the target word) could reveal the correct answer.

Method

Participants

$N = 32$ preschool-aged children ($M = 4.30$ years, $SD = 0.68$, 19 female) from the same pool as in previous experiments took part.

Stimuli

Visual stimuli matched previous experiments. Learning phrases were identical to those in Experiment 4, with each talker labeling the other talker's favorite. Test phrases matched Experiment 2, in which the liker's name (or pronouns *I* or *me*) occurred prior to the target word.

Procedure

This matched Experiments 2–4.

Results

Accuracy

Like previous experiments, accuracy exceeded chance (Fig. 3, far right) on both first-person (other-talker) trials ($M = .664$, $SD = .200$; $t(31) = 4.27$, $p = .0002$) and third-person (original-talker) trials ($M = .678$, $SD = .255$; $t(31) = 4.07$, $p = .0003$). An ANOVA assessed accuracy, with Trial Type and Block as within-participants factors. No effects were significant.

Visual fixations from word onset

Visual fixations diverged from chance soon after word onset (Fig. 10). This suggests that children did take advantage of liker cues early in the sentence, particularly for third-person trials, which appear to show a greater target advantage than first-person trials.

To assess these observations statistically, an ANOVA was computed on target advantage with Trial Type, Block, and Time Window as factors. Two participants were dropped for having less than 50% looks to referent pictures in the analyzed time span (results were similar with these participants included). The effect of Time Window reached significance ($F(1,29) = 9.29$, $p = .004$), with increasing target advantage as time increased. Trial Type approached significance ($F(1,29) = 3.43$, $p = .07$), with slightly more looks on third-person trials. Individually, *t*-tests indicated that while third-person trials showed significantly above-chance looks ($t(31) = 3.71$, $p = .0009$), first-person trials did not ($t(29) = 1.35$, $p = .19$). This difference, explaining the marginal effect of Trial Type, hints at a slight looking advantage for the third-person trials. No other effects or interactions approached significance. Individual *t*-tests compared the two trial types in each time window, and compared each trial type to chance in each window (all Bonferroni-corrected for five tests; see Fig. 10). These indicate that third-person trials exceeded chance in the 200–400 ms window, while first-person trial looks were beginning to reach significance (which did not survive Bonferroni correction) near the end of the time window. However, trial types never differed from each other. This pattern of results suggests that children do learn person-referent mappings when they are learning from third-person utterances.

Looks from sentence onset

As in Experiment 2, looks were realigned to sentence onset (Fig. 11) and analyzed to establish the exact time course of looks. One participant was dropped for having less than 50% looks to referent pictures in the analyzed time span (results were similar with this participant included). Unlike Experiment 2, no reliable differences were observed as a function of initial vs. non-initial liker (pro)-noun, so this factor was not considered. An ANOVA with Trial Type, Block, and Time Window (200–400, 400–600, 600–800, 800–1000, 1000–1200) was conducted on target advantage. The only effect approaching significance was

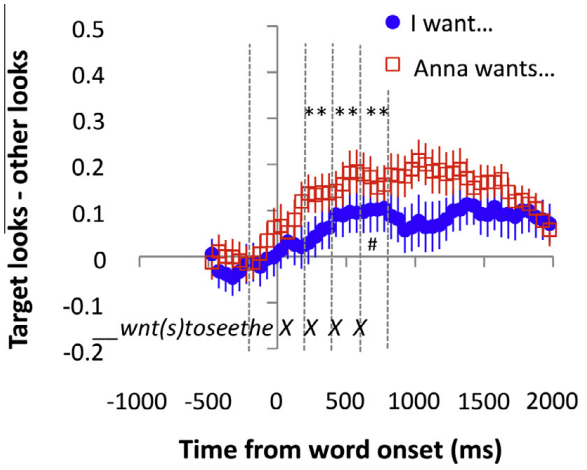

Fig. 10. Experiment 5 visual fixations. $^{\#}p = .05$, but not significant after Bonferroni correction. **Bonferroni-corrected $p < .01$. All *p*-values refer to difference from chance. Symbols above the hollow squares refer to hollow squares; those below filled circles refer to filled circles.

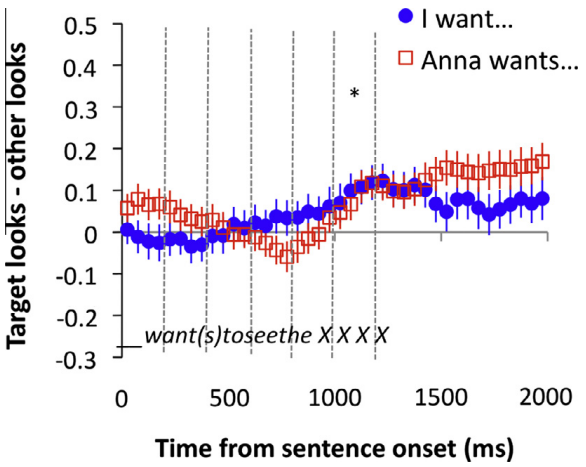

Fig. 11. Experiment 5, looks to pictures from sentence onset. *Bonferroni corrected $p < .05$, collapsed across trial types.

Time Window ($F(1,30) = 3.02$, $p = .09$), indicating a tendency for target advantage to increase over time. There was no effect of Trial Type, nor any significant interactions. Bonferroni-corrected *t*-tests on overall target looks (see Fig. 11) were conducted in each time window, indicating divergence from chance in the 1000–1200 ms time window only. Looking patterns on the whole appear less robust than those observed in Experiment 2.

Discussion

While looking patterns in Experiment 4 did not reflect learning of person-referent mappings, looking patterns in the current experiment showed learning of person-referent mappings: children reliably looked to the liker's favorite picture prior to word offset. This suggests that children *can* learn what each talker likes from the other talker's descriptions—that is, they learn and use person-referent mappings. This pattern of results implies that children's apparent failure to *use* person-referent mappings in Experiment 4 cannot stem from failure to *learn* person-referent mappings. Thus, we are left with only two explanations of Experiment 4: that there was either a conflict between person-referent mappings and voice-referent mappings, or a conflict between person-likes-referent mappings and person-talks-about-referent mappings.

Interestingly, even though children seemed to use third-person reference to decide which picture to look at, the effect of first-person pronouns was harder to interpret. First-person pronoun effects looked somewhat weaker than those seen in Experiment 2, which used the same set of test sentences. These weaker effects might stem from a conflict between either of the pairs of conflicting cues described above.

Table 3
Summary of results from all experiments.

	Cues present		Early target looks?
	Voice cues/ knower cues	Liker cues	
Experiment 1			
Talker-distinguished	√	(√)	Yes
Single-talker	.	.	No
Experiment 2			
First-person	√	√ (early)	Yes
Third-person	X	√ (early)	Yes
Experiment 3			
First-person	√	√ (late)	Yes
Third-person	X	√ (late)	Yes (but wrong way)
Experiment 4			
First-person	X	√ (late)	No
Third-person	√	X (late)	No
Experiment 5			
First-person	X	√ (early)	Yes? (marginal)
Third-person	√	X (early)	Yes

Note: √ = cue predicted correct response; (√) cue potentially predicted correct response but was not as emphasized as in other experiments;. = cue was absent; X = cue predicted incorrect response.

General discussion

At the outset, three questions were raised. First, do children use talker information to recognize words, and if so, how do they do so? Second, do children encode person-referent mappings even when learning new word-referent mappings? Third, how do children resolve conflict between voice cues (talker-specific words or voice-referent mappings), semantic cues (person-referent mappings), and phonological (word) cues? Five experiments (summarized in Table 3) addressed these questions.

The first question, and perhaps the largest question, was whether children use talker information during word learning. The answer is a qualified yes. Similar to related adult studies (Creel & Tumlin, 2011; Creel et al., 2008), children were no more accurate at learning words distinguished by a talker cue than words that were not (Experiment 1). Nor were children any *less* accurate when voice information mismatched the voice at training (Experiments 2–3, third-person trials; Experiments 4–5, first-person trials). However, visual fixation patterns across experiments suggest that children used talker information to recognize words *faster*. They looked more rapidly to referents associated with a particular individual's voice, either in the absence of other cues (Experiments 1 and 3), or when the speaker refers to herself/himself with a first-person pronoun (Experiment 2 and possibly Experiment 5, first-person trials).

Also under examination is *how* talker information was used. Using voice information to infer identity when hearing a first-person pronoun suggests that children encoded *person-referent mappings* and accessed those mappings via voice cues to identity. Children could also look to voice-matched referents because they have talker-specific word representations. This was supported somewhat by Experiment 1, wherein children did not show a talker-specific looking advantage until after word onset. Had they been using *voice-referent mappings*, they should have looked to voice-matched targets sooner, that is, prior to word onset. Of course, it could be the case that the sentences in Experiment 1 were too brief to reveal this. Alternatively, children in Experiment 1 could have been using *person-referent* mappings—the person speaking always talks about the pointy-looking creature, a more high-level, semantic use of talker information. This use of *person-talks-about-referent mappings* would be consistent with one account of Experiment 4, that children had learned *double* person-referent mappings which conflicted with each other: a person-likes-referent mapping and a person-talks-about-referent mapping.

However, there is still some ambiguity in the data between voice-referent and person-referent mappings. Specifically, children in Experiment 3 looked to voice-matched pictures in the temporary absence of information about who the liker was. Experiment 4 tried to determine whether this reflected a low-level *voice-referent mapping*—these voice acoustics cooccur with that picture—or a high-level *person-referent mapping*—this talker is going to ask for that picture. However, when Experiment 4 pitted these two cues against each other, there was little evidence of

either mapping. This did not stem from poor learning of person-likes-referent mappings (Experiment 5), suggesting that the pragmatics of someone talking only about someone *else*'s favorite thing did not impede learning. Thus, the most likely remaining possibilities are that children did learn person-referent mappings throughout, but in Experiments 4 and 5, they learned either a conflicting *voice-referent* mapping or an additional, conflicting person-talks-about-referent mapping. The latter possibility would suggest very high-level use of talker information. It would also be consistent with the somewhat larger effect magnitude in Experiment 2 (where person-likes-referent and person-talks-about-referent pointed in the same direction) relative to Experiment 5 (where the two mappings pointed in different directions). Finally, it may be consistent with Experiment 1, where talkers in the talker-distinguished condition did not espouse any partiality to a particular referent, but clearly had knowledge about one referent, allowing person-talks-about-referent mappings, with no conflicting person-likes-referent mappings.

The second question raised in the Introduction was whether children encode person-referent mappings even when learning new word-referent mappings. As addressed copiously above, the experiments taken together suggest that children learn person-referent mappings and word-referent mappings at the same time, as reflected in their use of cues to the liker (proper nouns or first-person pronouns). There is even a possibility that children may make multiple semantic mappings—though this is based partly on a null effect which could also be explained as a voice-referent mapping, and as such requires further investigation.

A final question was what children might do if the multiple cues they learned were in conflict. The experiments taken together suggest that children adeptly resolve conflict between these cues. First, Experiments 2 and 5 showed that children can use (pro)nominal reference to a talker, even though voice cues conflict, to identify the relevant individual's preferences. Second, in Experiment 3, eye movements on third-person trials clearly showed children using voice-referent (or person-talks-about-referent) mappings early in the sentence, but switched their looks to the target once the word was phonologically disambiguated. Finally, Experiment 3 (plus the single-talker condition of Experiment 1) suggests that children can use phonological information—the final, disambiguating phoneme of the target word—to begin directing fixations toward the correct referent. These patterns suggest that children make inferences or linguistic predictions from a variety of cues—person-referent mappings, lexical-phonological cues, perhaps voice cues—and that they readily recover when information later in the sentence conflicts.

How do listeners represent acoustic attributes of talker?

In the Introduction, a distinction was made between acoustic encodings of talker information (talker-specific word representations or voice-referent mappings) vs. semantic encodings of talker information (person-referent mappings). Evidence here pointed mainly to person-referent mappings, but other mappings cannot be ruled out. A relevant issue, both for the current study and the field in general, is what listeners do with talker-related components of the speech signal. Particularly, it is an open question whether children and adults represent acoustic attributes of talkers separately from acoustic–phonetic attributes of speech. On the one hand, it is trivially clear (and very important!) that children can *understand* unfamiliar voices (even unfamiliar accents, according to Schmale & Seidl, 2009, and Schmale et al., 2010; though see Nathan, Wells, & Donlan, 1998), so there must be some abstraction-like mechanism that allows children to recognize words despite talker variation, and talkers despite word variation. Nonetheless, other studies (e.g. Johnson, Strand, & D'Imperio, 1999; Mullennix et al., 1989; Niedzielski, 1999; Palmeri et al., 1993) make clear that talker differences are linked to phonetic differences. These studies, along with recent studies of language-familiarity effects on voice recognition (Bregman & Creel, 2014; Perrachione, Del Tufo, & Gabrieli, 2011), support the possibility of substantial overlap between speech representations and talker representations even in adults. Perhaps children and adults are simply better at selectively attending to word-relevant acoustic features than infants are (e.g. Houston & Jusczyk, 2000).

The current study, given that there was only partial evidence consistent with talker-specific word representations, sheds little light on developmental change in the degree of overlap or separation of speech vs. talker characteristics. As noted in the Introduction, there is some evidence that both infants (Houston & Jusczyk, 2000) and adults (e.g. Palmeri et al., 1993) encode talker information as an aspect of a word's form. From that perspective, children in the current age range might do so as well. However, the evidence for talker-specific word representations here is equally well-explained by person-referent or voice-referent mappings. For example, looks to the target picture on talker-distinguished trials in Experiment 1 are not actually differentiable from first-person talker-distinguished trials in Experiments 2 or 3, if measured from the earliest point where talker information is available (sentence onset; see Appendix A for differences in word onset time). This means that children in Experiment 1's talker-distinguished trials may have been doing something like what children in Experiments 2 and 3 were doing—assuming that the talker was going to ask for the picture they talked incessantly about.

Further, the paradigm used was deliberately simple, asking children to learn only two words at a time. This simplification set up the exact situation that led adults to use an apparent voice-referent mapping prior to word onset in Creel and Tumlin (2011): the voice perfectly predicts the target. It might be necessary to put children in a situation where voice does *not* predict the picture perfectly to see talker-specific-word effects in isolation. Future work should examine this possibility to establish whether there are developmental increases in selective attention to speech vs. talker characteristics.

Children's representations of talkers' mental states

The current study may shed light on children's representations of talkers' mental states. An interesting aspect of these results is that children seem fairly fluent at encoding person-referent mappings—the picture preferences of each talker—even though they are also encoding word-referent mappings. Recall that Creel (2012) found that preschool-aged children can keep in mind two talkers' preferred colors and can execute eye movements to pictures of those colors based on hearing that talker's voice. The current study extends Creel's (2012) finding to newly-learned words, which is impressive in that it means children are simultaneously encoding novel word-referent mappings as well as person-referent mappings. This finding also extends those of Borovsky and Creel (in press), who showed that 3–10-year-olds can use voice information to access familiar long term knowledge about talkers' roles and integrate that with sentence structure. The current study shows that children can also use voice information to access *newly-acquired* knowledge about a talker's preferences.

Implications for the development of language comprehension

The field of language development has been interested in how readily children integrate multiple sources of information. Earlier studies suggest that 5-year-olds have difficulty using pragmatic cues to constrain reference, and also have great difficulty revising initial erroneous sentence interpretations (Trueswell et al., 1999; see also Weighall, 2008), unlike adults (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The current results are highly distinct in asking how readily children can both encode and deploy novel mappings in sentence processing. Results suggest good encoding *and* deployment of newly-learned information, and deft revision of an erroneous initial interpretation. One implication of these different patterns is that learning person-referent mappings is particularly easy for preschool-aged children, while pragmatic cues based on visual scene membership, as in Trueswell et al. (1999), may be more subtle and require lengthier learning.

Preschool-aged children's apparent ease in storing person-referent mappings along with word-meaning mappings in turn implies that much of their language input may be conditioned on the source from a very early age. Earlier studies (Creel & Jimenez, 2012; Mann, Diamond, & Carey, 1979) suggest that children's encoding of voices may be somewhat coarse, perhaps limited to age and gender categories and native vs. non-native speaker status (as in Kinzler et al., 2007), though children may have better representations of familiar voices (e.g. Spence, Rollins, & Jerger, 2002). Nonetheless, the possibility that children condition their language input on a speaker's gender, age, and nationality would mean that language learning is quite context-dependent early in life. Additionally, children might supplement voice cues to identity with visual cues (face recognition) or contextual cues (I am in a grocery store vs. religious meeting vs. school). Those identity and situation cues, in combination with voice information, might contextualize the child's language input from a very early time point.

Conclusion

A series of experiments suggested that preschool-aged children simultaneously encode novel word-referent mappings and person-referent mappings. Talker information is clearly used in the service of person-referent mappings: when children heard "I want…", they appeared to use voice cues to access semantic information about the person associated with that voice. Other uses of talker information (talker-specific word representations, voice-referent mappings) cannot be ruled out. Further, children use these mappings flexibly to identify which talker's preferences were relevant in a particular sentence. Results imply that children encode multiple cues to meaning concurrently, and easily resolve conflicts between these cues, in contrast with previous studies of children's cue integration in language processing (Morton & Trehub, 2001; Trueswell et al., 1999).

Acknowledgments

Thanks to Adrienne Moore, Dolly Rojo, Annie Ditta, Emilie Seubert, and Nicolle Paullada for running participants, to Conor Frye and Annie Ditta for providing voices, and to Annie Ditta for extensive editing of sound files. Funding was provided by NSF CAREER Award BCS-1057080.

Appendix A. Acoustic properties of test sentences (SDs in parentheses)

See Fig. A1 and Table A1.

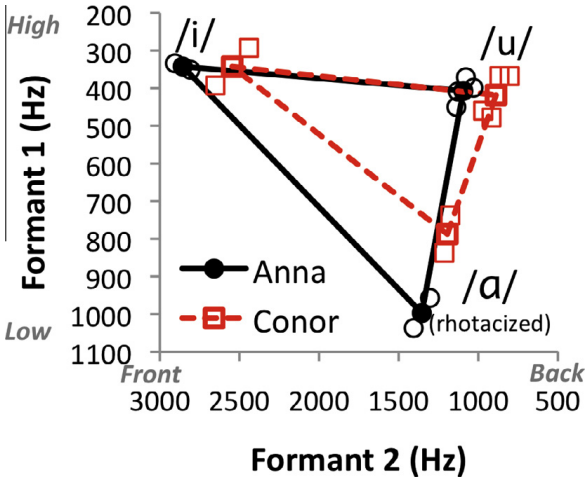

Fig. A1. Vowel triangles for the two talkers, illustrating higher F2 in the female talker ($t(7) = 6.53$, $p = .0003$). Unattached circles and squares represent individual tokens, attached circles and squares represent means. Vowels were measured for Experiment 2 learning phrases "The X is sooooo cool. Go, X!" The vowel /u/ was measured in the word "cool" (4 tokens per talker); a slightly rhotacized /ɑ/ was extracted from the final token of marv and mard (2 tokens per talker); /i/ was extracted from the final token of geeb and geege (2 tokens per talker).

Table A1
Timing and pitch characteristics of talkers.

	Sentence duration (ms)	Word duration (ms)	Word onset (ms)	Word POD (ms)[a]	Mean f0 (Hz)	f0 range[b]
Exp. 1						
Anna	1177 (237)	624 (65)	553 (231)	534 (40)	308 (40)	3.19 (0.38)
Conor	1115 (160) +	590 (76) +	524 (145)	411 (34) +	221 (22) ***	2.58 (0.80) **
Exps. 2 and 5						
Anna	1615 (276)	623 (77)	992 (270)	429 (42)	291 (22)	3.06 (0.49)
Conor	1479 (257) ***	591 (97) *	888 (234) ***	400 (41) **	209 (21) ***	2.50 (0.66) ***
Exps. 3 and 4						
Anna	1912 (309)	502 (111)	503 (126)	387 (69)	321 (32)	3.30 (0.47)
Conor	1778 (370)	494 (94)	471 (100) *	377 (66)	221 (23) ***	2.81 (0.39) ***

Note: Speech segments with creaky voice were not considered in measurements of pitch minima used in the f0 range calculation.
+ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .0001$.
[a] POD = point of disambiguation measured from word onset.
[b] As pitch is perceived on a log scale, pitch range is calculated as the ratio of highest pitch to lowest pitch, such that higher ratios indicate greater pitch variation.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635.

Borovsky, A., & Creel, S. C. (in press). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*.

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

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

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*, 85–95.

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

Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development, 83*, 2042–2056.

Creel, S. C. (2014). Impossible to _gnore: Word-form inconsistency slows preschool children's word-learning. *Language Learning and Development, 10*, 68–95.

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

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

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.

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition, 4*, 483–489.

Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory & Cognition, 5*, 658–665.

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.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. 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. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570–1582.

Jerger, S., Martin, R., & Pirozzolo, F. (1988). A developmental study of the auditory Stroop effect. *Brain and Language, 35*, 86–104.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359–384.

Kinzler, K. D., & Dejesus, J. M. (2013). Northern = smart and Southern = nice: The development of accent attitudes in the United States. *Quarterly Journal of Experimental Psychology, 66*, 1146–1158.

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

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition, 87*, 179–208.

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35*(3), 512–520.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*, 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263–285.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*, 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /1/: A first report for publication. *Journal of the Acoustical Society of America, 89*, 874–886.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 391–409.

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., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology, 40*, 255–265.

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

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365–378.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science, 13*, 407–420.

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

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62–85.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, speech production, and linguistic structure*. Washington: IOS Press.

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America, 102*, 3711–3722.

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., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science, 333*, 595.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

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.

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

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy, 15*, 650–662.

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.

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. 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.

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. F., & 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.