Arts	Arts Research & Publications

2018-01

When it is apt to adapt: Flexible reasoning guides children's use of talker identity and disfluency cues

Graham, Susan A; Thacker, Justine Marie; Chambers, Craig G.; Graham, Susan A.

Elsevier

When it is apt to adapt: Flexible reasoning guides children's use of talker identity and disfluency cues

Justine M. Thacker [a], Craig G. Chambers [b], Susan A. Graham [a],*

[a] Department of Psychology, University of Calgary, Calgary, Alberta T2N 1N4, Canada
[b] Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

ARTICLE INFO

ABSTRACT

An eye-tracking methodology was used to examine whether children flexibly engage two voice-based cues, talker identity and disfluency, during language processing. Across two experiments, 5-year-olds ($N = 58$) were introduced to two characters with distinct color preferences. These characters then used fluent or disfluent instructions to refer to an object in a display containing items bearing either talker-preferred or talker-dispreferred colors. As the utterance began to unfold, the 5-year-olds anticipated that talkers would refer to talker-preferred objects. When children then encountered a disfluency in the unfolding description, they reduced their expectation that a talker was about to refer to a preferred object. The talker preference-related predictions, but not the disfluency-related predictions, were attenuated during the second half of the experiment as evidence accrued that talkers referred to *dispreferred* objects with equal frequency. In Experiment 2, the equivocal nature of talkers' referencing was made more apparent by removing neutral filler trials, where objects' colors were not associated with talker preferences. In this case, children ceased making all talker-related predictions during the latter half of the experiment. Taken together, the results provide insights into children's use of talker-specific cues and demonstrate that flexible and adaptive forms of reasoning account for the ways in which children draw on paralinguistic information during real-time processing.

© 2017 Elsevier Inc. All rights reserved.

---

* Corresponding author.
    E-mail address: susan.graham@ucalgary.ca (S.A. Graham).

Introduction

Successful communication frequently relies on listeners' ability to attend to a variety of cues beyond the information conveyed by words alone. Consider, for example, a woman's utterance "I'll have my usual!" spoken by a regular customer in a restaurant. The meaning of this statement relies in part on the listener attending to the talker's identity (and memory for what her usual order is, e.g., a salad or pasta). These types of person-specific associations seem likely to interact with other kinds of paralinguistic cues found in spoken utterances. In the above example, imagine that the regular customer instead began her order with the following: "I'll have thee, uhh ...". Here, the talker's hesitation (marked by the filled pause "uhh") may suggest to the listener that she is, in fact, considering ordering something other than her usual order. Although there is considerable evidence that young children are adept at using different talker-produced cues to guide real-time referential interpretation, including talker identity and filled pauses, it is unclear the extent to which these abilities involve flexible forms of reasoning and whether these cues can be used simultaneously. The current study addressed these issues by investigating two questions. First, can 5-year-olds integrate two paralinguistic cues in the speech stream (talker identity and disfluency) to inform real-time referential predictions? Second, will 5-year-olds rapidly modify these predictions in response to counterevidence in the language they hear?

As background, previous research has demonstrated that children as young as 3 years can use talker identity cues to guide online language processing (e.g., Borovsky & Creel, 2014; Creel, 2012, 2014). For example, once preschoolers have learned that two distinct characters have a different preferred color, they demonstrate anticipatory looking to shapes bearing the talker's preferred color when listening to utterances produced by a given character. This is based entirely on cues to talker identity carried in the speech stream (Creel, 2012). Furthermore, children demonstrate flexible use of this cue. Specifically, whereas children use talker information to anticipate reference to talker-preferred objects when a talker speaks on her or his own behalf (e.g., Billy: "I want to see the square"), they will draw on knowledge of other individuals' preferences when the talker is speaking on their behalf (e.g., Billy: "Anna wants to see the circle"). It remains unclear, however, how referential expectations based on learned talker preferences are affected by evolving patterns as the discourse proceeds. That is, if the objects referred to by a talker in successive statements reflect a mix of talked-preferred and talked-dispreferred things, will child listeners dynamically adjust their earlier-formed expectations? If so, this would provide an additional source of evidence for the claim that children's use of talker identity cues relies on highly flexible forms of situation-specific reasoning.

Another point to consider is that research on children's use of talker identity cues has relied heavily on the link between preferences and desires, such that the talkers in these paradigms use desire statements (e.g., "I want ..."; Creel, 2012) to indicate that previously presented preference information is relevant to the current communicative context. Given that preference information can be relevant to referential intentions even when desire is not explicitly voiced, it is important to understand whether or how young children draw on a talker's identity to inform their interpretation of utterances that do not contain explicit statements of preference or desire. Hence, an additional goal of the current study was to examine whether children use recently presented preference information even when utterances are purely ostensive in nature such as simply directing listeners to "look" at an object in the display. If young children demonstrate anticipatory use of talker identity information in the absence of explicit desire statements, this would provide evidence for a subtler kind of sensitivity to talker preference information.

In the current study, we also considered talker effects alongside another important paralinguistic cue relevant to referential processing, namely *filled pause* disfluencies such as "um" and "uh". Filled pauses occur at a rate of approximately 2.56 per 100 words in adult speech and are associated with increased planning difficulty (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001). As such, filled pauses tend to occur in predictable locations in the speech stream such as in advance of noun phrases that refer to unfamiliar objects or to entities that have not yet been mentioned (Arnold & Tanenhaus, 2011). Although there is little research on listeners' reactions to spontaneous disfluencies, studies that

explore listeners' interpretations of filled pauses in experimental settings indicate that this systematic patterning of disfluencies does not fall on deaf ears. That is, both adult and child listeners actively use filled pauses as a cue to guide referential predictions, a process that has been deemed the *disfluency effect* (e.g., Arnold, Hudson Kam, & Tanenhaus, 2007; Arnold & Tanenhaus, 2011; Arnold, Tanenhaus, Altmann, & Fagnano, 2004). For example, disfluent referential descriptions lead children as young as 30 months to anticipate reference to unfamiliar discourse-new objects (Kidd, White, & Aslin, 2011; Orena & White, 2015) or to discourse-new objects that are familiar to both the listener and the talker (Owens & Graham, 2016). Thus, by their third year, children can use filled pauses to facilitate referential interpretation.

As with talker identity, there is emerging evidence to suggest that children's use of filled pauses depends on situation-specific factors. In the only study on this topic to date, Orena and White (2015) used a between-participants design whereby children were introduced to either a knowledge-able or forgetful talker who then produced fluent and disfluent utterances. Children exposed to the knowledgeable talker looked preferentially at novel discourse-new objects during disfluent utter-ances. In contrast, children in the forgetful talker condition did not display the same contrasts in antic-ipatory looking patterns, suggesting that 3½-year-olds can suspend their use of filled pauses when a talker has justified difficulty in naming objects.

In the current study, we explored the simultaneous use of disfluency and talker identity informa-tion to more fully understand the nature of children's situation-specific reasoning during language comprehension. Drawing on the methodology of Creel (2012), we introduced 5-year-olds to two char-acters, a male character and a female one, who were described as having opposite color preferences (i.e., blue and pink). In addition, the characters established an initial tendency to refer to objects that were of their preferred color. We then manipulated the fluency of the characters' referential descrip-tions on critical trials. On critical trials, a pair of pink and blue objects was presented on a large screen, and child participants heard either a fluent or disfluent utterance spoken by one of the two characters. On the assumption that preferred objects will be more salient to a given talker, and therefore more accessible to language production systems, it would follow that they would be more fluently described by that talker. Thus, when a talker is disfluent, the disfluency is likely to cue attention to an object that does not bear the talker's preferred color. Critically, the objects that would be associated with fluent versus disfluent descriptions should vary systematically according to the identity of the talker given the talkers' contrasting preferences. We reasoned that 5-year-olds' ability to generate and draw on these inferences should be reflected in gaze patterns to distinct objects on hearing disfluent versus flu-ent descriptions produced by the two talkers. Hence, a feature of the current experimental scenario is that children's potential use of disfluency cues does not rest on established associations between filled pauses and unfamiliar or hard-to-describe objects, as has been used in previous research on disfluen-cies (e.g., Kidd et al., 2011; Owens & Graham, 2016). Instead, the disfluency effect would arise more spontaneously from knowledge about talkers' preferences (which was presented just moments ear-lier) in combination with in-the-moment reasoning about what kinds of reference would hypotheti-cally lead to disfluency. In addition, talker specificity is not assessed by evaluating whether disfluency cues are neutralized for one talker (e.g., as in Orena & White, 2015); rather, it is assessed by evaluating whether the cues signal reference to distinct objects depending on the talker. If children can also modify their expectations about the informativeness of disfluencies in this new paradigm, this would provide additional evidence for children's adult-like appreciation that disfluencies can arise from a number of different sources in speech planning.

A second goal of the current study was to examine children's ability to dynamically update their assumptions based on the language patterns they encounter. This was inspired by recent work with adults showing that patterns in the language being heard, or the link between these patterns and par-ticular talkers, can shift aspects of semantic–pragmatic interpretation within the course of a single experiment. For example, adult listeners have been shown to adjust their interpretation of prosodic cues (Kurumada, Brown, & Tanenhaus, 2012), pragmatic inferences (Pogue, Kurumada, & Tanenhaus, 2016), and expectations for quantifier use (Yildirim, Degen, Tanenhaus, & Jaeger, 2016) in response to talker-specific biases and previous cue validity. In the current experiments, each talker referred to dispreferred- and preferred-color objects with equal frequency, and filled pauses occurred with equal frequency with each object type. In other words, these patterns do not, over the course of

the experiment, support a sustained association between disfluency and talker-dispreferred objects. If 5-year-olds can dynamically adjust their initial expectations based on the patterns they encounter, this would provide even further evidence for the flexible and context-sensitive nature of their use of paralinguistic cues.

The first experiment was designed to provide a test of children's ability to combine the two paralinguistic cues of interest (talker identity and disfluency) and to evaluate their ability to adapt their use of cues when talkers' utterances often provide counterevidence to children's initial assumptions. Experiment 2 was designed to replicate Experiment 1 in a situation where the relevant counterevidence was more evident.

Experiment 1

Method

Participants

Participants were recruited from a database of community families in a large metropolitan area in Canada. Data from 30 5-year-olds (16 boys; M_{age} = 5.19 years, SD = 0.17) were included in the final sample. An additional 5 children were tested but excluded from the analyses for the following reasons: technical error (n = 2), experimenter error (n = 1), and unsuccessful calibration on the eye tracker (n = 2). Children were primarily Caucasian with English as their primary spoken language, and 90% of parents reported having had at least some postsecondary education.

Apparatus and materials

The visual stimuli were presented on a 46-inch screen. As children listened to the recordings and viewed the images on the screen, their gaze position was tracked using a Tobii x50 system located on a table surface in front of the participants and underneath the display screen. The experiment was conducted using E-Prime with Tobii extensions. Specific areas of interest (AOIs), based on the location of display pictures, were set prior to data collection. Children's gaze positions were logged every 20 ms, and fixations were defined as looks to a location that lasted longer than 100 ms. We used Eye-gaze Language Integration Analysis (ELIA) software (Berman, Khu, Graham, & Graham, 2013) to align the gaze data with speech landmarks in the prerecorded utterances.

The test trials consisted of 56 cropped images of familiar objects, divided into 28 object pairs (see Appendix for a list of pairs). There were 16 critical trials involving object pairs consisting of one blue object (e.g., blue cat) and one pink object (e.g., pink drum). There was no repetition in the set of target and competitor object pairs. The remaining 12 trials were neutral filler trials involving objects whose colors were not associated with the stated speaker preferences (4 trials each with yellow–green object pairs, gray–brown object pairs, and red–orange object pairs).

Two adult native English speakers, one female and one male, recorded the auditory stimuli. We chose to use a male voice and a female voice because young children may have difficulty in distinguishing between same-gender speakers due to their greater acoustic similarity (Creel & Jimenez, 2012). Speakers first recorded an introduction for their character (see Fig. 1 for the script). Each introduction was played at the same time that a depiction of the corresponding character was present on the screen and established that each character had a preferred color (i.e., pink vs. blue). Preferred colors reflected common gender stereotypes (e.g., the female character preferred pink). After the introductions were completed, the depictions of the two characters were no longer displayed. Similar to Creel (2012), 8 alternating color-check trials followed the introductions. Color-check trials involved pairs of objects that were identical but differed in terms of color (e.g., blue sock/pink sock). The male speaker voice was used for instructions on 4 of the color-check trials ("Where is the blue one?"), and the female speaker voice was used for instructions on the other 4 trials ("Where is the pink one?"). These trials allowed us to verify that each participant could distinguish the colors, and reinforced children's knowledge of the characters' color preferences. For each of the 28 object pairs used in the main task (both critical and filler trials), four versions of an utterance instructing participants to look at one of the objects were recorded by crossing the two key variables of interest, voice and fluency: female

Introduction	
	Hi, my name is Sarah. I love things that are pink. I love my pink house, I love my pink toy, and I love my pink slippers. In fact, everything I own is pink. And most of all, I love to talk about things that are pink.
	Hi, my name is Adam. I love things that are blue. I love my blue house, I love my blue toy, and I love my blue slippers. In fact, everything I own is blue. And most of all, I love to talk about things that are blue.

Example Color Check Trial		
	Male Voice	Female Voice
	Where is the blue one?	*Where is the pink one?*

Example Test Trial			
		Male Voice	Female Voice
	Talker-Preferred Noun (Fluent)	*Look at the cat!*	*Look at the drum!*
	Talker-Dispreferred Noun (Fluent)	*Look at the drum!*	*Look at the cat!*
	Talker-Preferred Noun (Disfluent)	*Look at thee, uh, cat!*	*Look at thee, uh, drum!*
	Talker-Dispreferred Noun (Disfluent)	*Look at thee, uh, drum!*	*Look at thee, uh, cat!*

Fig. 1. Testing sequence with examples of visual and auditory materials. Images are a schematic of what was presented. Note that participants encountered a given object pair (e.g., cat/drum) only once per experiment and in one condition. However, all object pairs were cycled across each of the four conditions (female voice–fluent, female voice–disfluent, male voice–fluent, and male voice–disfluent) to create eight different orders.

voice–fluent, female voice–disfluent, male voice–fluent, and male voice–disfluent. Disfluent utterances were characterized by the pronunciation of the determiner "the" as "thee", followed by the filled pause "uh". The pronunciation "thee" (instead of "thuh") often precedes suspension of

speech in natural communicative exchanges and usually co-occurs with filled pauses (Fox Tree & Clark, 1997). Fluent utterances were recorded as "Look! Look at the X!" and disfluent utterances were recorded as "Look! Look at thee, uh, X!" Utterances were recorded in their entirety and were edited using Audacity, a multitrack audio editor. To ensure consistency across test trials, particular segments of the recorded utterances were standardized. Specifically, the initial "Look!" was excised from the same recording and then spliced into the fluent and disfluent versions. The average total length of fluent utterances was 7132 ms and the average length of disfluent utterances was 9251 ms.

Procedure

The experiment began with a gaze calibration program conducted using Tobii ClearView software. Accurate gaze calibration was required on at least three of the five test fixation points prior to initiating the experiment, with calibration achieved for all five fixation points for 90% of participants. When the experiment proper began, each child was presented with the character introductions, followed by 8 color-check trials, followed by a reintroduction to each character. Next, children viewed the 28 test trials that included 4 trials of each of the four critical trial types: 4 female voice–fluent trials, 4 female voice–disfluent trials, 4 male voice–fluent trials, and 4 male voice–disfluent trials. Critical and filler trials were randomly interspersed to create 8 trial orders. The left versus right location of the target object was also randomized across test trials.

Results and discussion

The measures of interest involve children's eye movement behavior at different time points during the unfolding instructions and how these patterns change during the first half versus the second half of the experiment. Because the talker identity and disfluency cues were relevant only for blue–pink trials, these critical trials were the focus of our analyses. The instructions "Look! Look at the X" or "Look! Look at thee, uh, X" were divided into two distinct intervals for analysis (see Fig. 2). For all intervals, a 200-ms margin was added to the boundary points to reflect the time lag in the execution of eye movements (Allopenna, Magnuson, & Tanenhaus, 1998; Matin, Shao, & Boff, 1993; Trueswell, 2008). To address whether children use talker information to form predictions about the intended object, we first examined fixations during the initial portion of the utterance, which we refer to as the *baseline interval.* This interval consists of the shortest duration of the interval corresponding to "Look! Look at ..." for both the fluent and disfluent utterances (1940 ms). We used the end point of this window (i.e., offset of "at") as the boundary point to ensure that no fluency information conveyed by the determiner was included in these analyses. We also explored how children's referential expectations were affected by the fluency of the unfolding description. To do so, we examined fixations in the portion of the utterance that included the determiner for both the fluent ("the") and disfluent ("thee, uh") utterances. The determiner interval was defined based on the duration of the shortest disfluent determiner ("thee, uh": 1700 ms), and its end point was aligned to just before the onset of the target noun to ensure that eye movements within this interval could not be driven by noun information. It is important to note that, because the duration of the determiner was by definition shorter in the condition with fluent utterances, the 1700-ms interval included 1480 ms of the baseline interval for fluent utterances. Although this entails reusing a portion of the previously analyzed time period (i.e., the baseline interval), the use of time intervals of equivalent duration is a necessary step to enable cross-condition comparisons where the potential to generate eye movements is fully equated in psychophysical terms. Lastly, an additional interval involving the disambiguating noun was analyzed separately simply to confirm that participants understood the instructions and were paying attention to the task. A one-sample *t* test indicated that the proportion of looks to the correct object was significantly greater than chance (.50) during this disambiguating interval (which was 1160 ms in duration, equivalent to the length of the longest noun): $M = .69$, $SD = .09$, $p < .004$.

Baseline interval

During this interval, sensitivity to talker identity would be reflected in anticipatory gaze shifts to the object bearing the current speaker's favorite color. To provide a measure of listeners' tendency to consider talker-preferred objects, we divided fixations to the talker-preferred object by the sum
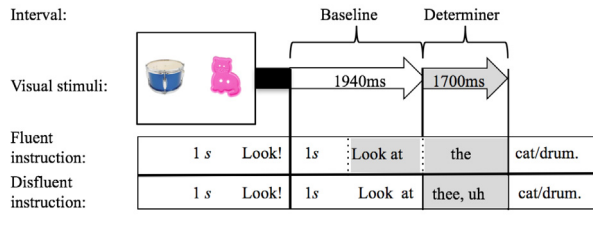
Fig. 2. The presentation of the object pairs was followed by instructions to look at one of the objects with either a fluent or disfluent description. Durations of the analysis intervals were 1940 ms for the baseline interval and 1700 ms for the determiner interval.

of fixations to both the talker-preferred and talker-dispreferred objects. By this measure, a value of 1 means that a listener fixated the talker-preferred object exclusively during a given interval, and a value of 0 means that a listener fixated the talker-dispreferred object exclusively. A value of 0.5 means that a listener spent an equal amount of time fixating both objects in the array. In addition, to determine whether children's initial predictions were altered over the course of the experiment, we compared the first half of the test trials (i.e., the first 8 critical trials) with the second half of the test trials (i.e., the latter 8 critical trials). Fig. 3 (left side) shows the average proportion of looks to talker-preferred objects during the baseline interval for the first and second halves of the experiment. Note that Fig. 3 shows data from both Experiments 1 and 2.

A two-level (Trials: first half vs. second half of experiment) repeated-measures analysis of variance (ANOVA) revealed a main effect of trials, $F(1, 29) = 4.755$, $p = .037$, $\eta_p^2 = .141$. Children's fixations to the talker-preferred object were significantly greater during the initial 8 critical trials ($M = .60$, $SD = .16$) than during the latter 8 critical trials ($M = .50$, $SD = .19$). Moreover, one-sample t tests confirmed that the proportion of fixations to talker-preferred objects was significantly greater than chance (.50) for the first 8 critical trials, $t(30) = 3.46$, $p = .002$, but not for the latter 8 critical trials ($p = .971$). Thus, on hearing the very first words of the utterance, and in advance of hearing any object label or cue except for the talker's voice, children's gaze behavior initially showed a prediction that talkers would refer to objects bearing their preferred colors. However, this expectation was neutralized as children learned that preferred and dispreferred objects, when present, were referred to in equal measure by each talker.

Determiner interval

Next, we analyzed children's gaze patterns during the determiner interval, which reveals how children react to disfluency cues. If these cues are spontaneously combined with talker-specific information to make referential predictions, children should expect that talker-preferred objects are more likely to be fluently described by the speaker and, conversely, that disfluencies are more likely to occur with descriptions for dispreferred objects. As before, we were also interested in how predictions were altered over the course of the experiment. Fig. 4 shows the average proportion of looks to talker-preferred objects during the determiner interval for both fluent and disfluent trial types. Note that Fig. 4 shows only data from Experiment 1.

The gaze proportion measure was the same as the one used in the baseline interval. Note that for these analyses, the n is equal to 28, reflecting the exclusion of 2 participants for whom zero looks were registered for all trials of a given condition during the determiner interval. A 2 (Fluency: fluent vs. disfluent) × 2 (Trials: first half vs. second half) repeated-measures ANOVA revealed a main effect of fluency only, $F(1, 27) = 4.821$, $p = .037$, $\eta_p^2 = .152$. There was no main effect of trials ($p = .188$), nor was there an interaction between trials and fluency ($p = .261$). Thus, children directed a greater proportion of looks to the talker-preferred object during fluent trials ($M = .55$, $SD = .15$) compared with during disfluent trials ($M = .45$, $SD = .18$). These results indicate that children were sensitive to the presence of
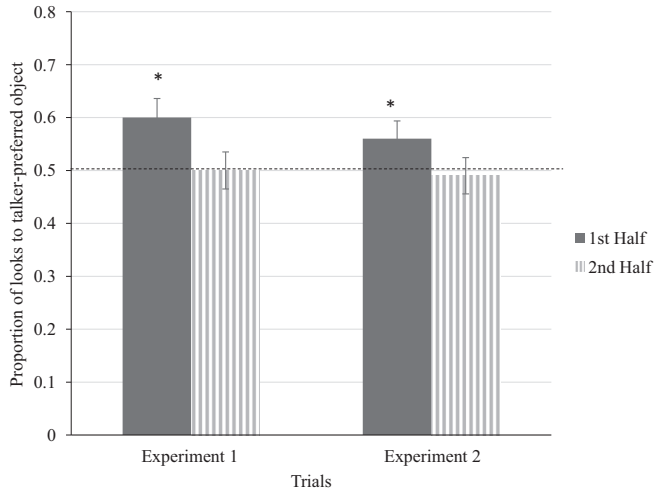
Fig. 3. Average proportion of looks to talker-preferred objects during the first and second halves of the experiment, during the baseline interval, for Experiments 1 and 2. In both experiments, the average proportion of children's looks to the talker-preferred object was significantly greater than chance (.50) for the first 8 critical trials (*ps* < .05) but not for the latter 8 critical trials (*ps* > .05). Error bars depict standard errors.
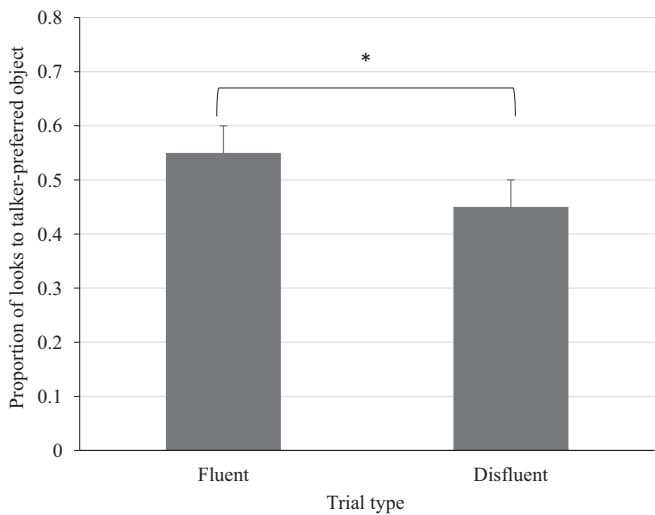

Fig. 4. Average proportion of looks to talker-preferred object during the determiner interval for both fluent and disfluent trial types in Experiment 1. Children showed a significantly higher proportion of looks to talker-preferred objects when encountering a fluent determiner compared with a disfluent determiner (*p* = .037). Error bars depict standard errors. A single asterisk represents significant differences at the p < .05 level.

disfluency and were biased to rapidly interpret filled pauses as signaling that the upcoming word would refer to something other than the contextually probable talker-preferred object.

The results of Experiment 1 revealed a robust sensitivity to talker identity during the first half of the experiment, which was then attenuated during the second half of the experiment. We discuss this finding further in the General Discussion. Although a similar pattern was not observed for the effect of filled pauses (i.e., an interaction such that children demonstrated a disfluency effect only for the first

half of the experiment), we conducted exploratory analyses separately for the first and second halves of the experiment. Consistent with our expectations, these results revealed a significant effect of fluency in only the first half of the experiment, $t(1, 29) = 2.063$, $p = .048$, such that children directed more looks to the talker-preferred object during fluent trials ($M = .60$, $SD = .22$) compared with disfluent trials ($M = .45$, $SD = .29$). The effect of fluency did not reach significance during the latter half of the experiment ($p = .422$). Thus, although a change in children's use of disfluency cues is numerically evident, the pattern was not strong enough to entail a significant Trials × Fluency interaction. In Experiment 2, we introduced a simple manipulation intended to increase the ability to detect children's sensitivity to the relevant patterns in the talkers' utterances.

Experiment 2

The purpose of Experiment 2 was twofold. First, we sought to replicate the findings of Experiment 1 and, second, we examined whether increasing the salience of disconfirming evidence would lead child listeners to more strongly attenuate their initial talker-specific referential predictions. Previous research has demonstrated that "prediction error" (i.e., recognizing that one's predictions were incorrect) leads to faster learning (e.g., Chang, Dell, & Bock, 2006; Ramscar, Dye, & McCauley, 2013). In the current case, the sooner listeners recognize erroneous predictions regarding talkers' reference to objects, the sooner they can recalibrate their expectations. Hence, in the current experiment, neutral trials were removed to increase the proportion of trials in which children encounter a prediction error. By removing these trials and keeping the rest of the experiment the same, the result was that we increased the percentage of disconfirming trials (i.e., the trials that fail to support a clear link between [dis]fluency and talker-[dis]preferred objects) from 29% of the entire set of trials (8 of 28) to 50% (8 of 16) while retaining the essential design elements of Experiment 1. We expected that the "condensed" counterevidence for an association between disfluency and reference to talker-dispreferred objects would result in a more detectable shift in children's use of disfluency cues.

Method

Participants
Data from 28 5-year-olds (13 boys; $M_{age} = 5.17$ years, $SD = 0.13$) were included in the final sample. Children were primarily Caucasian with English as their primary spoken language, and 90% of parents reported having at least some postsecondary education.

Materials and procedure
All visual stimuli were the same as the ones used in Experiment 1 with the exception that the 12 filler trials were not presented.
The procedure was the same as the one used in Experiment 1.

Results and discussion

As in Experiment 1, the focus of the current analyses was on children's eye movement behavior at different time points during the unfolding instructions and how these patterns change during the first half versus the second half of the experiment. The instructions "Look! Look at the X" and "Look! Look at thee, uh, X" were divided into the same two analysis intervals (baseline and determiner) used in Experiment 1.

Baseline interval
Recall that, during this interval, sensitivity to talker identity would be reflected in anticipatory gaze shifts to the object bearing the current speaker's favorite color. The gaze proportion measure was the same as the one used in Experiment 1. To determine whether children's initial predictions were altered over the course of the experiment, we compared the first half of the test trials (i.e., the first 8 critical trials) with the second half of the test trials (i.e., the latter 8 critical trials). Fig. 3 shows

the average proportion of looks to talker-preferred objects during the baseline interval for the first and second halves of the experiment.

A two-level (Trials: first half vs. second half of experiment) repeated-measures ANOVA revealed a main effect of trials, $F(1, 27) = 5.336$, $p = .029$, $\eta_p^2 = .165$, such that children's fixations to the talker-preferred object were significantly greater during the initial 8 critical trials ($M = .56$, $SD = .13$) than during the latter 8 critical trials ($M = .49$, $SD = .14$). Moreover, the proportion of fixations to talker-preferred objects was significantly greater than chance (.50) for the first 8 critical trials, $t(27) = 2.399$, $p = .024$, but not during the latter 8 critical trials ($p = .624$). Thus, on hearing the very first words of the utterance, and in advance of hearing any object label or cue except for the speaker's voice, children's gaze behavior showed an initial prediction for the direction of the utterance based on recently presented gender-stereotyped preference information. As in Experiment 1, however, the expectation that speakers would refer to talker-preferred objects was extinguished as children learned that speakers referred to preferred and dispreferred objects in equal measure.

Determiner interval

As before, this interval was used to measure the effect of the fluency manipulation. Recall that here, if children can use talker-specific information to make referential predictions, they should expect that talker-preferred objects are more likely to be fluently described by the speaker and, conversely, that disfluencies are more likely to occur with descriptions for dispreferred objects. Fig. 5 shows the average proportion of looks to talker preferred objects during the determiner interval for both fluent and disfluent trial types.

The gaze proportion measure was the same as the one used in Experiment 1. A 2 (Fluency: fluent vs. disfluent) × 2 (Trials: first half vs. second half of experiment) repeated-measures ANOVA revealed a Fluency × Trials interaction, $F(1, 27) = 7.116$, $p = .013$, $\eta_p^2 = .209$. During the first 8 critical trials, children directed a greater proportion of fixations to the talker-preferred object during fluent trials ($M = .48$, $SD = .18$) compared with disfluent trials ($M = .37$, $SD = .17$), $t(27) = 2.696$, $p = .012$. Children's proportion of fixations to the talker-preferred object was significantly below chance (.50) during disfluent trials, $t(27) = 4.066$, $p < .001$, indicating that children were sensitive to the presence of disfluency and were biased to rapidly interpret filled pauses as signaling that the upcoming word would refer to something *other* than the contextually probable talker-preferred object. Children's proportion of fixations to the talker-preferred object was not significantly above chance during fluent trials ($p = .58$). In contrast, during the latter 8 critical trials, the proportion of fixations to the talker-preferred object did not differ during fluent trials ($M = .45$, $SD = .18$) compared with disfluent trials ($M = .55$, $SD = .23$), $p = .085$. Moreover, children's proportion of fixations to the talker-preferred object during both fluent and disfluent trials did not differ from chance (.50), $ps > .129$. Hence, over the course of the experiment, child listeners dynamically adjusted their initial referential predictions.[1] This can be viewed as a rational response to the accumulating evidence, which failed to support a clear link between (dis)fluency and talker-(dis)preferred objects.

As in Experiment 1, an initial sensitivity to talker identity was observed during the first half of the experiment that was attenuated during the second half. That is, as a talker began to speak, children's gaze behavior showed an initial prediction for the direction of the utterance based on each talker's preferred colors. However, the expectation that talkers would refer to objects of a particular color was neutralized as children learned that preferred and dispreferred objects were referred to in equal measure by each talker. At the point when the fluent or disfluent determiner was encountered, children initially demonstrated an ability to use talker identity cues to guide their interpretation of a hesitation disfluency. This effect was also diminished during the latter half of the experiment due to the explicit counterevidence for children's initial spontaneous hypothesis for the basis of a hesitation disfluency. In contrast to Experiment 1, the more concentrated nature of counterevidence in Experiment

---

[1] To rule out the possibility that fatigue over the course of the experiment accounted for the neutralization of expectations observed in Experiment 2, we calculated target-looking scores during the noun region for the first and second halves of the experiment. Target-looking during the first half ($M = .75$) and target-looking during the second half ($M = .79$) were not significantly different, $p = .218$, indicating that fatigue does not explain the pattern of results.
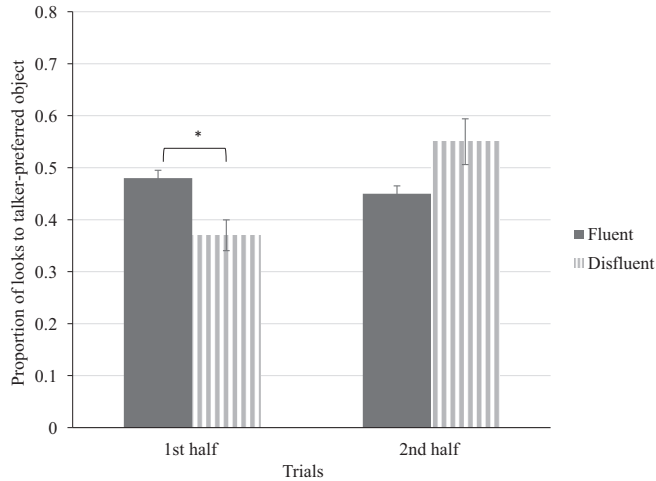
Fig. 5. Average proportion of looks to talker-preferred object during the first and second halves of Experiment 2, during the determiner interval, for both fluent and disfluent trial types. Children showed a significantly higher proportion of looks to talker-preferred objects when encountering a fluent determiner compared with a disfluent determiner during the first half of the experiment (p = .012). During the second half of the experiment, children did not show a significantly higher proportion of looks to talker-preferred objects during disfluent utterances compared with fluent utterances (p = .085). Error bars depict standard errors. A single asterisk represents significant differences at the p < .05 level.

2 led to a statistically robust disfluency by trials interaction. Thus, it appears that children will readily adapt their concurrent use of multiple paralinguistic cues in response to situation-specific evidence.

General discussion

The goal of the current study was to explore 5-year-olds' use of two paralinguistic cues, talker identity and filled pauses, to guide referential expectations. Specifically, we investigated whether children would engage both cues in the same utterance and whether sensitivity to these cues was flexible and responsive to distributional patterns in talkers' utterances.

Across two experiments, there was consistent evidence that children used recently presented information about talkers' preferences in conjunction with disfluencies to form anticipatory referential predictions. With respect to talker identity, we found robust sensitivity to this cue during the first half of the experiment. As an utterance began to unfold, children predicted that talkers were likely to refer to objects bearing their preferred color. This was the case even though the utterances were entirely ostensive in nature and simply directed listeners to "look" at an object in the display (and did not, e.g., request or otherwise express a desire for a particular object, as was the case in some previous work such as Creel, 2012). This anticipatory use of talker identity information, in the absence of explicit preference or desire statements, demonstrates that voice cues alone can, in some circumstances, influence children's referential predictions. As the utterance unfolded further, children spontaneously used knowledge of a talker's preferences to guide their interpretation of a hesitation disfluency in the unfolding utterance; when the disfluency was encountered, children reduced their expectation that the talker was about to refer to a preferred object. Taken together, these results document children's ability to use these two talker-produced cues in concert and also broaden our conceptualization of the "disfluency effect" (e.g., Arnold et al., 2007). More specifically, the results demonstrate that disfluency attributions can involve factors beyond the difficulty associated with planning descriptions for unfamiliar or as yet unmentioned objects.

A related objective of the current study was to explore children's ability to shift their use of paralinguistic cues in response to patterns in speech. In both experiments, talkers referred to objects bearing

a preferred or dispreferred color equally often across test trials and with equal numbers of fluent and disfluent descriptions. As a result, children's initial assumption that disfluency should accompany reference to dispreferred objects proved to be inconsistent with the emerging evidence as the successive test trials were encountered. From an adult perspective, this ought to lead rational listeners to "rethink" their original assumption, such that different expectations would be evident at the end of the experiment compared with the beginning. Our results clearly show that children are capable of recalibrating expectations in this same way. In Experiment 1, children reduced their reliance on talker identity, as evidenced by the absence of talker identity effects during the latter half of the experiment. In Experiment 2, there was a strong attenuation of *all* talker-specific referential predictions during the latter half of the experiment. These differing results can be attributed to the effect of the interspersed filler trials, which involved neutral colors that were not known to be preferred or dispreferred by talkers. The inclusion of filler trials in Experiment 1 had the effect of extending the spacing between disconfirming trials and reducing the proportion of trials containing disconfirming evidence. We suggest that this reduces the salience of disconfirming trials, which in turn slows the rate at which the initially assumed links are extinguished. With the removal of the filler trials in Experiment 2, the evidence that there was not in fact a clear link between (dis)fluency and talker (dis)preferred objects was correspondingly stronger, and as a result children were faster to adjust their referential expectations. Thus, although children can spontaneously use recently presented information about a talker's preferences (reinforced by gender stereotypes) in conjunction with disfluency cues to guide online predictions, they readily update their assumptions in response to disconfirming evidence, hence increasing the validity of their expectations. This ability to continuously reevaluate various associations is crucial given the dynamic and variable nature of real-world spoken language behavior. Although, as mentioned earlier, there is growing evidence that adults demonstrate pragmatic adaptation of this type within the course of a single experiment (e.g., Kurumada et al., 2012; Pogue et al., 2016; Yildirim et al., 2016; Yoon & Brown-Schmidt, 2014), this is, to our knowledge, the first set of studies to demonstrate a similar phenomenon in the real-time processing behavior of young children.

Comparisons with the findings from previous work are informative for understanding the *ease* with which children "update" the validity of speech-based cues in response to accumulated evidence. Earlier studies examining the nature of the disfluency effect in children (e.g., Kidd et al., 2011; Orena and White, 2015; Owens and Graham, 2016) also included disconfirming trials but did not report changes in the strength of children's predictions over the course of the experiment. For example, Kidd et al. (2011) used a 16-trial paradigm where disfluencies were equally likely to precede familiar and novel targets, thereby preventing children from learning any relation between disfluencies and target familiarity over the course of the experiment. Results indicated that children as young as 30 months looked more to the novel object during a 2-s window of analysis situated in advance of target word onset when there was a disfluency present (i.e., disfluent trials). Although these authors did not explore differences in the disfluency effect at the beginning versus the end of the experiment, the results (averaged across all trials) suggest a comparatively strong disfluency effect, unlike the eventually extinguished effect observed in the current study. These differing patterns likely highlight meaningful differences in the strength of the link between filled pauses and referential descriptions in different contexts. In previous studies (e.g., Kidd et al., 2011; Owens and Graham, 2016), the disfluency effect hinged on a link that children have likely forged based on a broad range of samples (i.e., the link between disfluent descriptions and given/new discourse status or with hard-to-describe or unfamiliar referents). In contrast, in the current experiments, children's initial sensitivity to disfluency cues was built on their understanding of objects the two talkers would likely want to mention, based on their stated preferences. It seems likely that this link would be comparatively fragile at the outset and that the talker-specific disfluency effects would be more likely to shift in response to disconfirming patterns in the language being heard.

Lastly, the results of the current study highlight the flexible ways in which children can use two paralinguistic cues, talker identity and disfluency, to guide referential expectations. The 5-year-olds demonstrated the ability not only to integrate both of these cues when processing a single utterance but also to adjust their cue-based expectations in response to talkers' referential patterns. What remains unclear is the precise extent to which the observed patterns rest on implicitly learned associations versus forms of pragmatic reasoning (see Arnold et al., 2007). Although an association-based

account is not incompatible with the results, we favor the explanation that the observed patterns reflect types of social reasoning, such that children in the current study were actively recruiting their knowledge of talkers' preferences and gender stereotypes alongside an abstract understanding that disfluencies signal a processing delay. We believe that this account provides a better overall fit with the available evidence, which includes evidence from Orena and White (2015) highlighting the role of speaker-specific inferences.

In summary, our results demonstrate preschoolers' impressive ability to adapt their concurrent use of two paralinguistic cues over time in response to situation-specific evidence. One promising direction for future research would be to examine children's flexibility in cases where talker cues become increasingly predictive of referential intent. In the research to date, including the current study, this flexibility has been explored by testing for the neutralization of cue use (e.g., Orena & White, 2015). Thus, it remains unclear whether preschool-aged children would demonstrate similar online flexibility when the patterns they encounter should serve to strengthen the use of paralinguistic cues during language processing rather than weaken them. Such a result would provide additional evidence for children's ability to rapidly and efficiently adjust referential predictions in response to the patterns children encounter in talkers' speech.

Acknowledgments

Appendix A. List of object pairs

Phase	Object 1	Object 2
Color-check	pink backpack	blue backpack
	blue book	pink book
	pink Band Aid	blue Band Aid
	pink flip-flop	blue flip-flop
	pink piggy bank	blue piggy bank
	blue block	pink block
	blue lunchbox	pink lunchbox
	pink hat	blue hat
Test	blue/pink towel	pink/blue bear
	blue/pink drum	pink/blue ball
	blue/pink car	pink/blue boat
	blue/pink scooter	pink/blue guitar
	blue/pink flower	pink/blue lion
	blue/pink comb	pink/blue bottle
	blue/pink box	pink/blue cat
	blue/pink key	pink/blue button
	green cup	yellow balloon
	green sock	yellow duck
	green dinosaur	yellow bucket
	green lamp	yellow horse
	brown dog	gray bell
	brown crayon	gray scissors
	brown scarf	gray boot

Appendix A (*continued*)

Phase	Object 1	Object 2
	brown chair	gray fork
	red hat	orange star
	red book	orange shirt
	red shovel	orange fish
	red mitten	orange hammer

References

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

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say *thee uh* you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science, 15*, 578–582.

Arnold, J. E., & Tanenhaus, M. K. (2011). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference* (pp. 197–217). Cambridge, MA: MIT Press.

Berman, J. M., Khu, M., Graham, I., & Graham, S. A. (2013). ELIA: A software application for integrating spoken language and eye movements. *Behavior Research Methods, 45*, 646–655.

Borovsky, A., & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*, 1600–1613.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123–147.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272.

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

Creel, S. C. (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language, 73*, 81–98.

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

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151–167.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science, 14*, 925–934.

Kurumada, C., Brown, M., and Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In N. Miyake (Ed.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 647–652). Austin TX: Cognitive Science Society.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Attention, Perception, & Psychophysics, 53*, 372–380.

Orena, A. J., & White, K. S. (2015). I forgot what that's called! Children's on-line processing of disfluencies depends on speaker knowledge. *Child Development, 86*, 1701–1709.

Owens, S. J., & Graham, S. A. (2016). Thee, uhh disfluency effect in preschoolers: A cue to discourse status. *British Journal of Developmental Psychology, 34*, 388–401.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology, 6.* https://doi.org/10.3389/fpsyg.2015.02035.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language, 89*, 760–793.

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 73–96). Amsterdam: John Benjamins.

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language, 87*, 128–143.

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 919–937.