# The many meanings of "um": Children draw rich, contextual social inferences after hearing speech disfluencies

Benjamin C. Morris[1], Daniel Yurovsky[1,2], & Alex Shaw[1]

[1] University of Chicago

[2] Carnegie Mellon University

## Author Note

Correspondence concerning this article should be addressed to Benjamin C. Morris, Department of Psychology, University of Chicago, 5848 S University Ave, Chicago, IL 60637. E-mail: benmorris@uchicago.edu

# Abstract

Timing is a crucial pillar of human communication (Stivers et al., 2009), and as such, disruptions to normative timing are marked and meaningful. These disruptions are often marked by disfluencies in one's speech (e.g., "um") that signal difficulty in planning or processing. By age 2, children produce disfluencies in their own speech, and draw referential predicitions after hearing a speaker produce a disfluency (Kidd, White, & Aslin, 2011). Adult listeners make a variety of social inferences from delay markers, including speaker knowledge (Brennan & Williams, 1995), and speaker desires (Roberts, Margutti, & Takano, 2011).


*Keywords:* communication, speech disfluency, mental-state inference, hesitation, uncertainty, processing time

Word count: X

# The many meanings of "um": Children draw rich, contextual social inferences after hearing speech disfluencies

Imagine you ask someone on the street, "Where is the nearest train station?" and they reply "It's um... that way." While plausibly correct, the stranger's reply is slow and marked, indicating uncertainty about accuracy of their statement. Despite getting a relevant answer, you yourself may still feel uncertain, perhaps choosing to confirm the direction by consulting another stranger or checking your smartphone. In conversation, *how* something is said may be as meaningful as *what* is said.

In the example above, the stranger seems uncertain because they spoke disfluently, pausing in the middle of their utterance. When it comes to how something should be said, language is profoundly structured by an expectation of timeliness. Conversations across at least 10 typologically diverse languages are structured around remarkably short and consistent silences between speakers (Stivers et al., 2009). Speakers of a given language show nuanced judgments for what consitutes a reasonable, or unreasonable, delay during a conversation, though the duration of that acceptable delay varies across languages (Stivers et al., 2009). Thus, conversation is powerfully structured to avoid delays when possible.

Producing language in a timely manner may often conflict with demands on processing, memory, and many other factors liable to cause delays. Imagine trying to produce the answer to a difficult question or remember the plot of a movie you saw years ago– starting or finishing your utterance in a timely manner might be quite difficult. When such delays are unavoidable, speakers may often mark them with disfluencies (e.g., filled pauses "uh" and "um").

Conversation is rife with these small disruptions to otherwise smooth, alternating language exchange (Fox Tree, 1995; Shriberg, 1996). These disfluencies are constant; some data sources estimate that 6 disfluencies occur every 100 words in adult conversations (Fox

Tree, 1995). While ubiquitous in everyday language use, disfluencies are commonly held by laypeople as errors in production with little meaning. Disfluencies typically arise from difficulties in planning one's production and signal an upcoming delay in speech, for example when retrieving an infrequent word or feeling uncertain of the answer (Clark & Fox Tree, 2002; Smith & Clark, 1993). Far from being random; disfluencies occur in predictable forms and locations in an utterance (Clark & Fox Tree, 2002).

Children likely have abundant and systematic experience hearing disfluencies in the language around them (Shriberg, 1996). Young children also have first-hand experience with disfluencies in their own productions– prevalent from at least age 2 (Casillas, 2014). 22-month-old infants distinguish between fluent and disfluent adult-directed speech, suggesting that children are sensitive to disfluencies in the second year of life (Soderstrom & Morgan, 2007).

Still more impressive, by their second birthday, toddlers are able to monitor these cues online and predict that a disfluent speaker will likely refer to a novel object that is new to the discourse, rather than a familiar one that has already been discussed (Kidd et al., 2011). Later in development this predicition may be tied to a speaker's knonwledge: 3 1/2 year-olds make the same prediction when pre-trained on a reliable speaker, but show no such prediction when pre-trained on a forgetful speaker who fails to remember the names familiar objects during training (Orena & White, 2015). Thus, from an early age, children reason about speech disfluencies and use them to guide referential inferences.

While disfluencies are clearly powerful cues for predicting language, our account emphasizes that they should also provide rich evidence for drawing social inferences. Disfluencies in speech may often reflect underlying processing time demands, perhaps due to delays in lexical search time or conflict in choosing one's utterance. To the extent that listeners are sensitive to this principle of processing time (Clark, 1996), disfluencies could serve as powerfully broad cues to the processes going on in other people's minds.

Indeed, there is some evidence that adult listeners use disfluencies to draw a range of social inferences, such as inferring a speaker's knowledge. When a speaker accurately, but disfluently answers a trivia question, adult listeners infer that person is less knowledgable about that piece of information, compared with an accurate answer that is speedy (Brennan & Williams, 1995). Interestingly, there is even some evidence that when speakers produced non-answers (e.g., "I don't know" or "I can't remember"), disfluency leads adults to infer greater knowledgeability– perhaps the disfluency derives from attempting to retrieve the answer before failing (Brennan & Williams, 1995). These results suggest adults may be reasoning about the underlying mental process that delays speech in powerfully flexible ways.

Adult listeners use delays to infer a speaker's desires. When responding to more socially-relevant questions (e.g., "Can you give me a ride?"), a speaker who is slower to respond is seen as less willing to accept the request (Roberts et al., 2006; Roberts et al., 2011). Similarly, when agreeing with another's past statement (e.g., "The flyers look good."), delays are seen as indicating less agreement (Roberts et al., 2011). These speaker judements are even titrated by the degree of delay– the longer the delay, the less willing a speaker seems (Roberts et al., 2011). Holding constant what a speaker says, the speed and fluency with which a speaker says it licenses broad, contextual inferences about their mental-states.

While no work has directly examined the development of these inferences, a small literature examines children abilities to reason about timing more broadly. In the physical domain, young children believe an agent who succesfully builds a tower faster than another agent building the same tower is better at building– though when the difficulty of the tower is also varied children have more trouble inferring relative competence (Leonard, Bennett-Pierre, Gweon, 2019). By age 7, children seem to use response time on a difficult problem to infer the likely underlying mental process, such that a quick answer is judged as a retrieved memory and a longer response time is judged as a in-the-moment solution (Richardson & Keil, 2020). Timing is an everyday force that even young children are likely very familiar with, and these studies suggest it may be a useful cue in the development of social reasoning.

In the present studies, we test how and when young children use speech disfluencies to draw social inferences about a speaker's knowledge and preferences across a range of conversation contexts. We show that children infer an accurate, but disfluent speaker is less knowledgable (Experiments 1a and 2a), but only in certain conversational contexts (Experiments 1b and 2a). Then in a new conversational context, we demonstrate that children are similarly able to use disfluency to infer a speaker's relative preference (Experiment 3). Ongoing work further examines the richness of children's preference inferences, such as inferring relative dispreference.

## Experiment 1a

When answering a question (e.g., "What is 5 times 7"), disfluencies indicate a delay in searching for an answer ("Um… 35"), which leads adults to infer a speaker is less knowledgeable even if they offer accurate information (Brennan & Williams, 1995). A large literature demonstrates that children by age 4 use language *accuracy* to judge knowledgeability– for example, children prefer to learn from a previously accurate speaker rather than someone who mislabels or fails to label common objects (e.g., Koenig & Harris, 2005). The present study extends such work to ask how chilidren consider not just what a speaker says, but how they say it. Thus, in all of the present studies, accuracy is held constant.

Relevant prior accuracy information may rarely be available to the developing learner, or they might have limited knowledge on which to judge past accuracy. Moreover, in many situations, a speaker's accuracy alone may convey little about their underlying knowledge of the given domain (e.g., when discussing a relatively simple topic). Here, a better cue might be the speaker's ease and fluency in discussing the topic at hand. In a pilot study, we demonstrate that children ages 4-9 similarly use speaker disfluencies to infer knowledgeability.

## Methods

**Participants.** 22 children (ages 4-9, mean age = 7.54, 14 girls) were recruited at a local science museum. Due to timing constraints for collecting data in a museum collection, sample demographics beyond participant gender were not collected.

**Procedure.** Children were presented with an animated story on an iPad about two speakers and two familiar stuffed animals (a cat and a dog). In the story, each speaker is asked "What is this animal called?" about the animal closest to them, and the experimenter reads their replies. In Experiment 1a, one speaker fluently labels their animal ("This one is a cat") and the other speaker disfluenty labels the other animal ("This one is a… uh… dog"). Across participants, speaker order was counterbalanced, such that half the children heard the disfluency first. Children were then asked a domain-wide knowledge question: "Who do you think knows more about animals– this person or this person?" while the experimenter pointed to each speaker. If children failed to choose one of the speakers (e.g., saying "both") or failed to respond within 5 seconds, the experimenter repeated the question one time. Children completed only a single trial.

## Results

Asked which speaker knows more about animals, children were signficantly more likely to select the fluent speaker: 0.86% of children (19/22) reported that the fluent speaker knows more (binom $p < 0.001$).


*Figure 1*. Stimuli used for the labelling condition in Experiment 2.

## Experiment 1b

Is this because children think disfluencies always signal a lack of knowledge? Many disfluencies in naturalistic speech are related to discourse history, speech rate, or interlocutor familiarity and may not indicate knowledgeability (Clark & Fox Tree, 2002; Shriberg, 1996). Thus, adults do not always infer that disfluencies indicate incompetence, for example when answering a question with a non-answer (e.g., "Uh… I don't know"). In this context, adults may even treat disfluency as a signal of greater knowledge– the hesitation implied by a disfluency suggests the speaker feels like they might know the answer and attempted to search for it (Brennan & Williams, 1995).

In a pilot study, we demonstrate that children ages 4-9 do not simply use speaker disfluencies to assume a lack of knowledge, but instead flexibly refrain from making this inference depending on the context. Specifically, we show that children do not selectively infer knowledge when both speakers produce non-answers.

### Methods

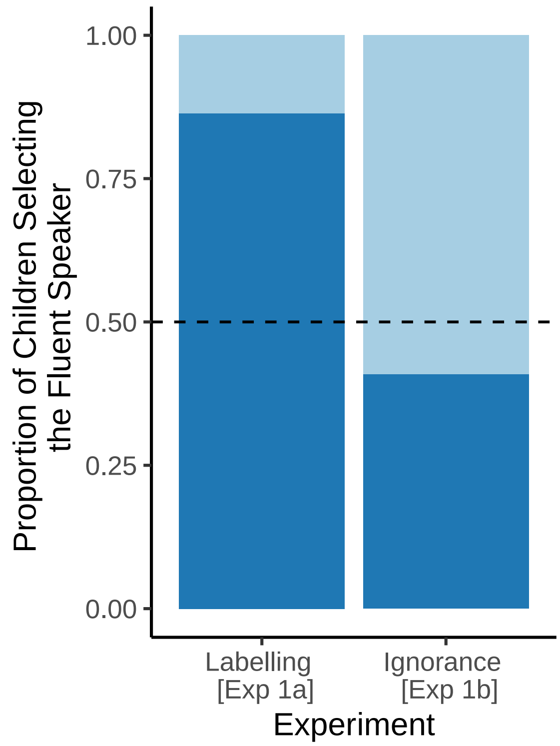


*Figure 2*. Results from pilot Experiments 1a and 1b, plotting participant repsonses when asked "Who knows more about animals?" The dashed line indicates chance responding.

**Participants.** 22 participants (ages 4-9) were recruited at a local science museum. Data from 1 additional participant was excluded due to experiment error.

**Procedure.** In Experiment 1b, children were presented with an animated story on an iPad about two speakers and two stuffed animals selected to be less familiar to children (an aardvark and an anteater). As in Experiment 1a, each speaker is asked "What is this animal called?" about the animal closest to them, and the experimenter reads their replies. In Experiment 1b, both speakers claim ignorance and produce a non-answer, but one does so fluently ("I don't know") and the other does so disflueuntly ("Uh... I don't know"). Across participants, speaker order was counterbalanced, such that half the children heard the disfluency first. Children were then asked to select the speaker they think "knows more about animals." If children failed to choose one of the speakers (e.g., by saying "both") or failed to respond within 5 seconds, the experimenter repeated the question one time.

Children completed only a single trial.

## Results

Asked which speaker knows more about animals, children were at chance choosing the fluent and disfluent speaker with similar liklihood: 0.41% of children (9/22) reported that the fluent speaker knows more (binom *ns*).

# Experiment 2a

In two separate pilot studies, we saw preliminary evidence that children ages 4-9 selectively infer competence based on delay markers, even when both speakers are accurate, but also understand such inferences are not always licensed. Building off these pilot results, Experiment 2a was a pre-registered experiment to replicate our pilot effects in a larger sample and examine potential age-related change. Participants were randomly assinged to either the Labelling or Ignorance condition. We also added an additional trial (with new animals and speakers) to allow us to collect more judgments from each child.

Based on our pilot data, previous adult work (Brennan & Williams, 1995), we expected that even 4-5 year old children would infer an accurate but disfluent speaker would be less knowledgeable (labelling condition). We also predicted that children would make no such inference in the ignorance condition, and that older children may even infer that the disfluent speaker knows more in this condition, as suggested by the adult work (Brennan & Williams, 1995).

## Methods

**Participants.** We pre-registered a planned sample of 120 children, 60 children in each of the two conditions (labelling vs. ignorance). For each condition, we planned to collect data from 20 children in each of 3 pre-deterimend age-groups: 4-5 years-old, 6-7 years-old, and 8-9 years-old. Data were collected at a local science museum. Due to the suspension of human subjects research following the COVID-19 pandemic, we were unable to complete data collection with the 8-9 year-old sample. As of March 2020, we have data from 101 children (40 children ages 4-5, 40 children ages 6-7, and 21 children ages 8-9).

**Procedure.** The procedure was nearly identical to Experiments 1a and 1b. Two characters in a story are asked to label one of the two familiar stuffed animal on the table in front of them. In the Labelling condition, both speakers correctly label the animal in front of them, but one does so disfluently as in Experiment 1a (see Figure 1). In the Ignorance condition, both speakers produce non-answers, but one does so disfluently– saying, "Uh… I don't know" as in Experiment 1b.

In Experiment 2a, we matched the stuffed animals in the story across conditions, and also added an additional trial with two new animals and two new speakers. For the anmials, we selected familiar animals with no readily available alternative labels (as opposed to "dog", which may acceptably be labeled "puppy"). In trial 1 (for both the Labelling and Ignorance conditions) the speakers were asked about a tiger and a zebra. In trial 2, the speakers were asked about a cow and a pig. Across participants, we counterbalanced the speaker order (whether the first speaker was fluent or disfluent) and the location of the two animals, yielding 4 counterbalanced orders. Note that this also counterbalances which animal the disfluency is paired with across participants. In the second trial, speaker order was always the reverse of the first trial.

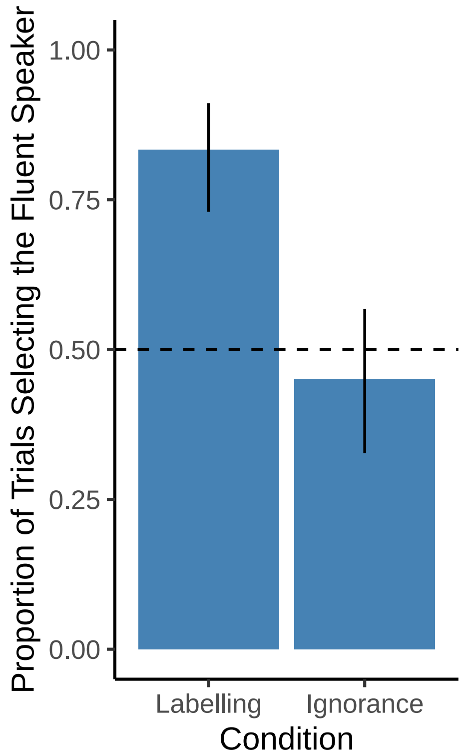


*Figure 3*. Overall results from Experiment 2a, with bootstrapped 95% confidence intervals (black lines). The dashed line indicates chance responding.

## Results

Overall, our results are consistent with our initial pilot results from Experiments 1a and 1b (see Figure 3). Looking at the proportion of trials that children select the fluent speaker as more knowledgable, we see that children choose the fluent speaker as more knowledgeable in the Labelling condition (mean proportion of trials = 0.83) signficantly more than chance ($t(50) = = 7.29$, $p < .001$ ). In the ignorance condition, children's responding did not differ from chance (mean proportion of trials = 0.45), suggesting they were not relaibly selecting either speaker ($t(49) = = -0.84$, $p$ .403).

To compare choices across conditions, we used a mixed effects logistic regression prediciting speaker choice by condition and age (continuous), with random effects of subject and trial number. There was a signficant negative effect of condition, such that children were

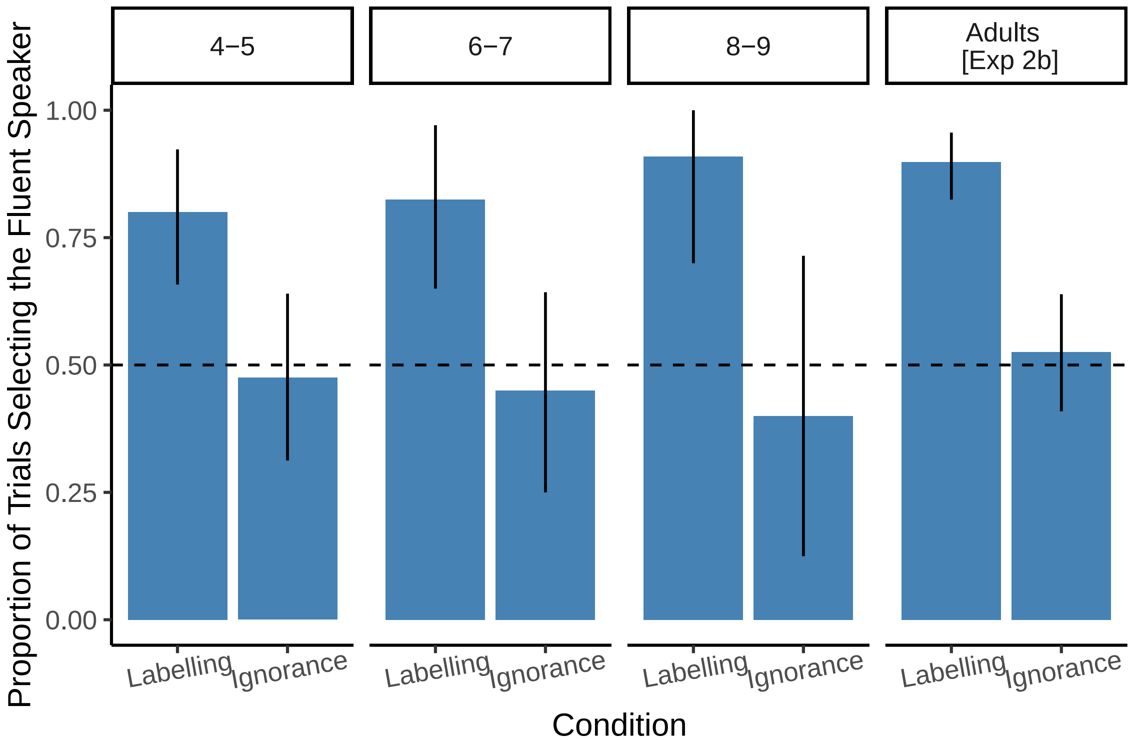


*Figure 4*. Age-binned results from Experiment 2a, with bootstrapped 95% confidence intervals (black lines). The dashed line indicates chance responding.

significantly less likely to choose the fluent speaker as more knowledgeable in the ignorance condition, compared with the Labelling condition ($\beta$ = -0.36, $p < .001$). The effect of age was not signficant ($\beta$ = 0.02, $p$ = .308).

Even the youngest children in our sample show this pattern (see Figure 4). Looking only at the 4-5 year-olds, children selected fluent speaker as more knowledgeable in the Labelling condition (mean proportion of trials = 0.80) signficantly more than chance ($t(19)$‘ = 4.49, $p < .001$). In the ignorance condition, 4-5 year old children's responding did not differ from chance (mean proportion of trials = 0.48), suggesting they were not relaibly selecting either speaker ($t(19)$ = -0.29, $p$ = .772).

## Discussion

In Experiment 2a, children make robust, selective knowledge inferences, judging an accurate but disfluent speaker to be less knowledgeable than an accurate and fluent speaker. Supporting our initial pilot results, we also see evidence that even 4-5 year old children are consistently making this inference, with no evidence of developmental change. These results suggest that children are tracking speech disfluencies and using them to make social inferences about another person's knowledge from a young age, as adults do (Brennan & Williams, 1995). This work adds to a small number of demonstrations that young children's knowledge inferences are senstive to not just what you answer (i.e. accuracy), but also how you answer (e.g., Aboody, Huey, & Jara-Ettinger, 2018).

Children show no reliable speaker preference in the ignorance condition. While a null result, this finding helps rule out the idea that children are heuristically tracking disfluencies and always ascribing incompetence. This finding also suggests that our results cannot be explained by low-level auditory fluency. There is evidence that preschool age children are sensitive to auditory fluency cues, judging speakers whose utterances were inflected with white noise to be less competent than matched speakers who spoke without such background noise (Bernard, Proust, & Clément, 2014). However, such an effect cannot explain these results as the ignorance condition contains an identical amount of disfluency, yet children make no such knowledge inferences.

Previous work in adults suggests that they differentiate fluent and disfluent non-answers, judging disfluent non-answers to indicate greater knowledgeability than fluent non-answers (Brennan & Williams, 1995); however, in these data we saw no evidence that kids made such a differentiation. Thus, in our next experiment we collected adult judgements about these very same stimuli.

## Experiment 2b

To determine the adult-like pattern of responses in this task, we collected knowledge judgments from a sample of adults online. While past studies have addressed similar questions in adult samples (Brennan & Williams, 1995), our task differs along a number of dimensions and so a separate baseline of adult responses is required. Additionally, while Brennan & Williams (1995) reported that adult listeners inferred that disfluent non-answers indicated more knowledge, we saw no evidence that children were making this inference in Experiment 2a. In Experiment 2b, we can ask whether adults systematically prefer the disfluent speaker in ignorance condition, as the results from Brennan & Williams (1995) might suggest.

### Methods

**Participants.** 120 participants were recruited from Amazon Mechanical Turk, with 60 participants in each of the two conditions (labelling vs. ignorance). Participants were paid a small reward in exchange for completing the study. Two participants had incomplete data and were excluded from the final sample.

**Procedure.** Experiment 2b was an online adaptation of Experiment 2a. In Experiment 2a, the experimenter read the story and speaker's utterances aloud to children. Adult participants were instead asked to read through survey slide-style qualtrics, and then select which speaker would know more about animals. Otherwise, the stimuli and trial structure were exactly the same. As in Exeriment 2a, particpants completed two trials.

### Results

Adult responses closely mirrored children's responses (see Figure 4). Looking at the proportion of trials that adults select the fluent speaker as more knowledgable, we see that adults choose the fluent speaker as more knowledgeable in the Labelling condition signficantly more often than chance. In the ignorance condition, adults did not reliably select either speaker.

## Experiment 3

Our account holds that disfluencies can be interpreted by listeners as domain general signs of processing time or conflict. As such, disfluencies should generate inferences in a variety of domains, beyond knowledgeablility. We next ask how children use speech disfluencies to infer an agent's preferences and dispreferences.

In Experiment 3, we change the question under discussion so that each character is asked which of the two animals they like the best. Disfluency here may indicate a speaker is experiencing conflict between the two options, delaying their response time, while speed may indicate a strong and decisve preference (similar to the degree of agreement inference made by adults in Roberts et al., 2011). When two speakers both state the same preference, but one does so disfluently, we predicted that children would infer they had a relatively weaker preference. We further expected that this situation might license an additional inference about dispreference: namely, when asked who likes the unlabelled more (and thus dispreferred item), we expected that children would infer the disfluent speaker may have a relatively stronger preference. We expected that the dispreference inference might be difficult for our youngest children, based on some evidence that infants seem better able to represent preference rather than dispreference (Feiman, Carey, & Cushman, 2015).

Data for Experiment 3 were conducted remotely over Zoom, as described more below.

To establish that our paradigm sucessfully translated to remote testing, we also included an additonal knowledge trial in Experiment 3, attempting to replicate our findings from Experiment 2a.

## Methods

**Participants.** We pre-registered a planned sample of 60 children to run in Experiment 3. Again, we planned to collect data from 20 children in each of 3 pre-deterimend age-groups: 4-5 years-old, 6-7 years-old, and 8-9 years-old. These data were collected online via Zoom with a live experimentor. Particpanting families were largely recruited via a participant database of Chicagoland families who have previously participated in in-person research studies. These data are still being collected, with a current sample of 56 children (16 children ages 4-5, 20 children ages 6-7, and 20 children ages 8-9).

**Procedure.** In Experiment 3, the general procedure was largely matched to the previous experiments. We made some changes to facillitate the shift to remote data collection, such as changing the response format (described below). The central change of interest was to shift the question under discussion.

***Preference Trial.*** In Experiment 3, two characters are introduced one at a time, and each character is given an identifiable color by which they are referred (e.g., "the blue person"). This change facillitated children's ability to respond verbally when asked to select a character, rather than pointing to a character. In Experiment 3, each speaker enters the scene independently, and is alone when the question is asked. Viewing the tiger and the zebra, each speaker is asked, "Which of these animals is your favorite? Which one do you like the best?" Both speakers independently state a preference for the same animal, but one does so disfluently, saying e.g., "Um… the tiger is my favorite." At the end of the trial both speakers are then brought back and children are reminded what each speaker said. This ensures that both speakers are onscreen when the target questions are asked. Across participants, we counterbalanced whether the first or second speaker is the disfluent speaker, and which animal is preferred.

Children are then asked three questions. The preference question: “Who do you think likes the tiger more?”, the distractor question: “Who do you think is better at playing basketball?”, and the dispreference question: “Who do you think likes the zebra more?” The order of the preference and dispreference questions was counterbalanced across participants. The distractor question was always asked second to ensure children were not merely switching responses across questions. Children completed one preference trial.

***Knowledge Trial.*** After switching to remote testing due to the onset of the Covid-19 pandemic, we wanted to ensure that our paradigm successfully translated from in-person adminsitration. As an initial check, we included an additional trial always after the key preference trial described above. We made a few additional alterations to our paradigm, thus this additional trial cannot serve as a true replication.

The Knolwedge trial is structured exactly like the preference trial described above. As in Experiments 1 and 2, the characters in the story are asked, “What is this animal called?” but in Experiment 3 both are asked to name the same animal independently. Both speakers label the animal accurately, but one does so disfluently (e.g., saying “This one is a… um… pig.”).

Children are then asked three questions. The labelled knowledge question: “Who do you think knows more about [pigs]?”, the distractor question: “Who do you think is better at playing soccer?”, and the unlabelled knowledge question: “Who do you think knos more about [cows]?” The order of the labelled and unlabelled knowledge questions was counterbalanced across participants. The distractor question was always asked second to ensure children were not merely switching responses across questions. Children completed one knowledge trial.

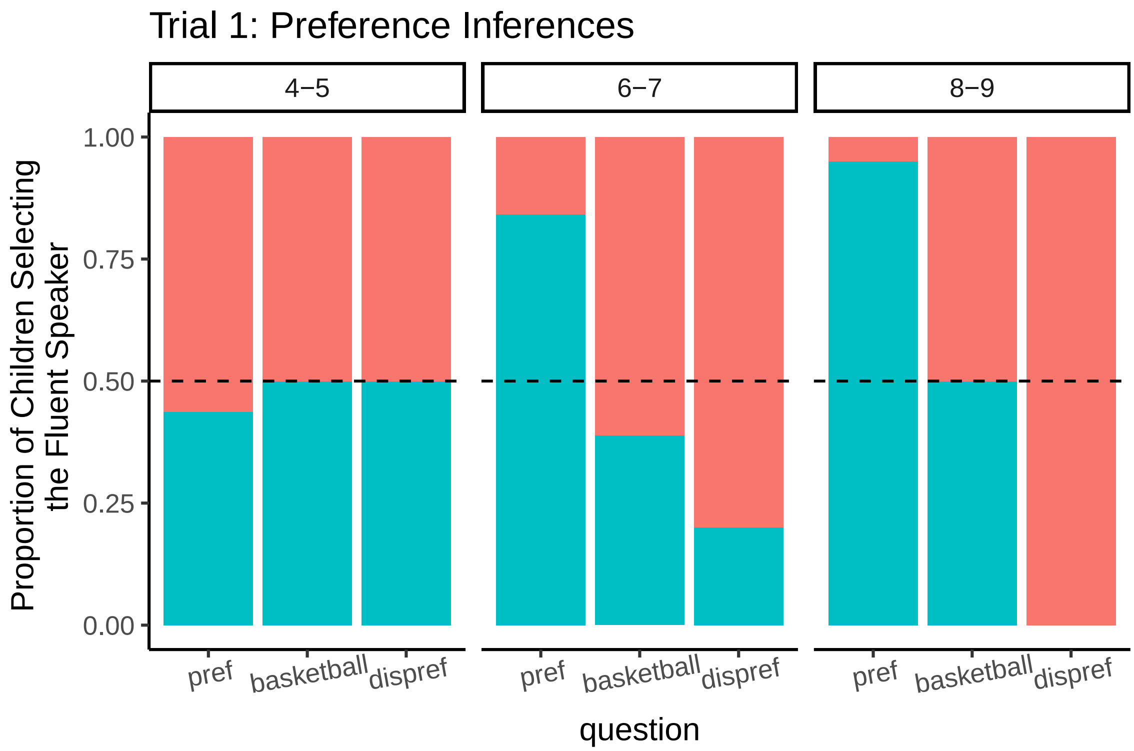


*Figure 5*. Preference trial results from Experiment 3.

## Results

Note that data colleciton remains ongoing for the youngest age group (4-5 year-olds, current n = 16/20) in Experiment 3, so we will only report overall effects at this time.

**Preference.** We first analyze children's judgements from trial 1 (the preference trial). When asked which speaker likes the preferred animal more, children were signficantly more likely to select the fluent speaker. When asked which speaker likes the dispreferred animal more, children were signficantly less likely to select the fluent speaker. When asked the distractor question, children showed no robust preference for one speaker over another.

**Knowledge.** Next, we analyze children's judgements from trial 2 (the knowledge trial). We see preliminary evidence that the updated paradigm administered over Zoom

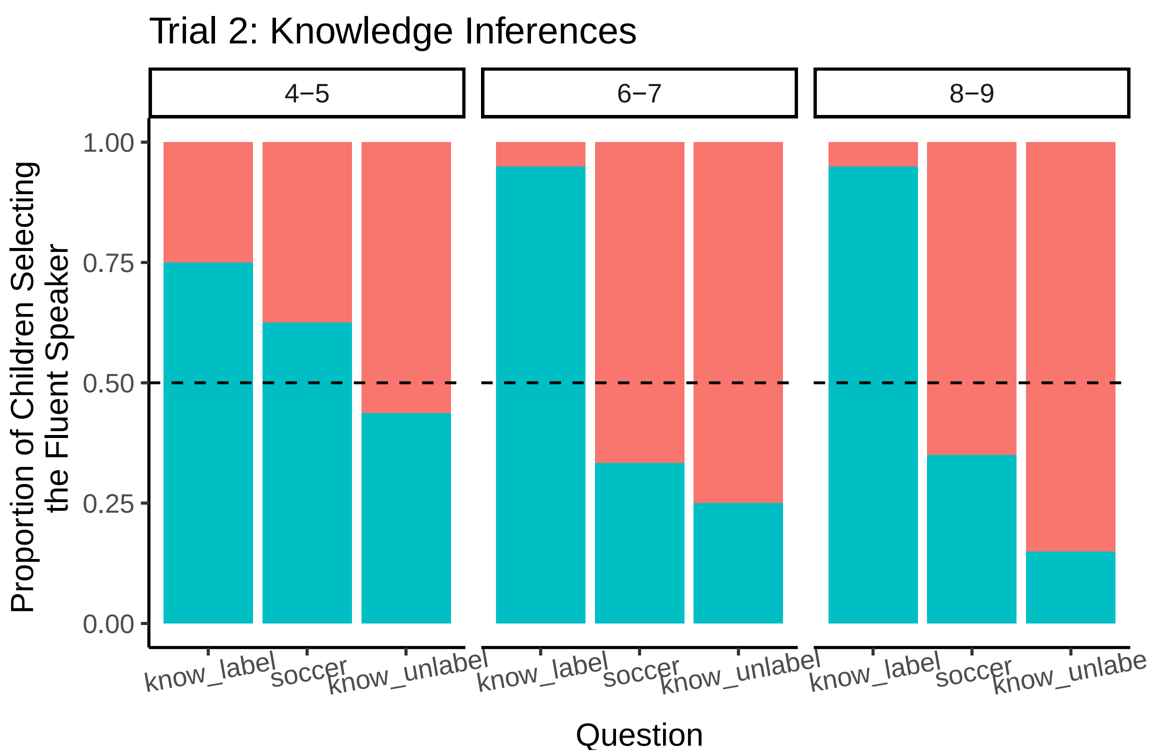


*Figure 6*. Knowledge trial results from Experiment 3.

provides a conceptual replication of our in-person results from Experiment 2. Asked which speaker knows more about the labelled animal species, children were significantly more likely to select the fluent speaker than the disfluent speaker. This suggests that our paradigm successfully adapted to the Zoom administration. Additionally, when asked the distractor question, children showed no robust preference for one speaker over another.

We expected that when asked who knows more about the unlabelled animal, children would be at chance or even favor the fluent speaker. However, children were signficantly more likely to select the disfluent speaker.

## Discussion

Overall, we find support for the idea that children use disfluencies not just to make knowledge inferences, but to make preference inferences as well, consistent with an account where disfluencies serve to signal general delays in processing with contextual intepretation.

While our data collection is still ongoing with the youngest age-group (4-5 year-olds), the current data suggest these youngest children may not be drawing reliable preference inferences. This would be an interesting contrast to our findings from Experiment 2a, which are replicated here, that 4-5 year-olds are able to infer knoweldgeability in a closley matched task. One interesting possiblity is that connection between disfluency and knowledge emerges before it is connected to other domains like preference, at least in this task.

The data on children's inference of dispreference are more difficult to interpret. While children do reliably select the disfluent speaker as having a stronger preference for the unlabelled object, we see a similar, unpredicted pattern for children's knowledge inferences about the unlabelled animal. That is, while children made domain-wide competence inferences in Experiment 2a, and reliably report that the fluent speaker knows more about the animal they mentioned, they seem to make the opposite inference when asked about another, unlabelled animal. These data make the dispreference measure difficult to interpret, since this inference may be driven by order effects or task demands. In this experiment, the preference trial was our focus, thus the knowledge trial was always run afterwards. It is possible that the pattern of responses evident in trial 1 carried over to trial 2 and affected children's knowledge judgements as well. Ongoing work examines the effect of order on this inference.

## General Discussion

Across 3 studies, children draw inferences about another agent's mental states based solely on the disfluencies in their speech, and these inferences are flexible and context-sensitive. We see consistent evidence that even young children infer that a disfluent speaker might be less knowledgeable. By age 6, children similarly use disfluencies to infer the relative preferences of a speaker. In sum, these findings suggest that disfluencies may serve as powerful cues to a speaker's mental processes, with broad applications.

These studies add to the rich literature on children's ability to infer knowledgeability and engage in selective social learning (e.g., Koenig & Harris, 2005). In our studies, children make selective competence judgements, even while holding speaker accuracy constant. These results parallel recent demonstrations that young children are able to use accuracy irrelevant information in selective social learning paradigms (Aboody et al., 2018; Corriveau, Kurkul & Arunachalam, 2016). In our study, 4-5 year old children (as young as we tested) made these contextual knowledge inferences, at rates similar to adults.

We see preliminary evidence that these knowledge inferences may emerge earlier in development than preference inferences in this task. Across both Experiment 2 and 3, children age 4-5 seem to infer a disfluent speaker might be less knowledgable, but do not extend this inference to the preference domain in Experiment 3. These preliminary results present the intruiging possiblity that children may more easily connect fluency and knowledge. Future work is needed to test the robustness of this developmental change.

While adults derive a range of social inferences based on speech timing and fluency (Brennan & Williams, 1995; Roberts et al., 2011), these inferences are likely built on general inferences about processing delays and their richness comes from contextualized interpretation. That is, these delays often acquire social meaning only after integrating information from the conversational context– such as inferring that someone who says "sure" slowly is less willing to go along with previous speaker's request. It is possible that the preference inference in our experiments required more contextual information than the knowledge inference and was more difficult for young children as a result.

Children are not merely responding heuristically to the percieved confidence of a speaker, but instead flexibly and contextually employing these inferences. Children as young as 2 are sensitive to an agent's confidence (e.g., facial expression, gesture, etc.) and selectively imitate more confident agents (Birch, Akmal, & Frampton, 2010). Our results cannot be explained by tracking confidence, as a confidence account should predict systematic choices in the ignorance conditions where children and adults show no speaker preference. We also see that children will select the disfluent speaker as having a stronger dispreference, though this effect is difficult to interpret in these data.

Disfluencies are powerful cues for inference– available in-the-moment and informative regardless of whether the child has the relevant knowledge to judge a speaker's statement While these cues are doubtless helpful in language processing, this work extends such findings to suggest that these cues additionally serve a range of social inferences in early childhood. Disfluencies track processing difficulties, and thus can provide a useful window into an agent's mental processes, beyond the content of what is said. As young children learn about the social world, tracking speech disfluencies and processing delays broadly may provide a rich dataset for drawing social inferences.

## Acknowledgement

The authors are grateful to Cassidy Wilson and Madison Collins for their assistance in data collection.

## References

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, *14*(4), 925–934.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of american english, italian, and japanese. *Discourse Processes*, *48*(5), 331–354.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … others. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, *14*(4), 925–934.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of american english, italian, and japanese. *Discourse Processes*, *48*(5), 331–354.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … others. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.