

# Toddlers Actively Sample from Reliable and Unreliable Speakers

**Jess Mankewitz (mankewitz@wisc.edu)**

Department of Psychology, University of Wisconsin-Madison

**Robert D. Hawkins (rdhawkins@wisc.edu)**

Department of Psychology, University of Wisconsin-Madison

**Jenny Saffran (jsaffran@wisc.edu)**

Department of Psychology, University of Wisconsin-Madison

## Abstract

Toddlers are sensitive to the reliability of speakers in their environment (Koenig & Woodward, 2010). While previous work suggests that children prefer labels from reliable speakers, it remains unclear how these social representations guide lower-level information-seeking processes that affect speaker preferences. The current study introduces a gaze-contingent eye-tracking paradigm to investigate how children engage in sampling during word learning. Toddlers (22-24m) view videos of two speakers labeling familiar objects; one speaker provides reliable labels and the other speaker provides unreliable labels. Toddlers then sample novel labels from the speakers using a gaze-contingent interface: only the speaker they are fixating on provides a novel label. Preliminary data ( $N = 18$ ) suggests that participants prefer to sample first from the reliable speaker over the unreliable speaker. However, there is little difference in overall sampling preferences. Our findings suggest that toddlers can assess speaker reliability, but remain open to exploring information from unreliable speakers.

**Keywords:** Active Learning, Eye-tracking, Reliability, Language Acquisition

## Introduction

Children are active members of their learning environment. Before they produce their first words, curious infants explore and collect information about the world around them (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Kidd & Hayden, 2015; Chu & Schulz, 2020; Lapidow & Walker, 2020). In isolation, they can explore the world through their own senses and experiences (Ruff, 1986). However, as social learners, infants can also leverage their knowledge of other agents to elicit additional information they could not acquire alone (Begus & Southgate, 2018; Waismeyer & Meltzoff, 2017). By pointing, vocalizing, and directing their gaze, they can direct the attention of people in their environment toward objects of interest and receive information about their functions, properties, or, often, labels.

Social cues are thus particularly crucial when learning language (Hembacher, deMayo, & Frank, 2020; Lee & Lew-Williams, 2023). For example, 6-month-old infants are more likely to follow the gaze of an informant after an ostensive cue like infant-directed speech or direct eye gaze (Senju & Csibra, 2008). Moreover, because language use is flexible and varies from context to context, children must attend not only to the mappings between potential referents and labels but also to qualities of the *informant*. In particular, a robust research literature has shown that young children are sensitive

to the reliability, or correctness, of speakers in their environment (Harris, Koenig, Corriveau, & Jaswal, 2018; Koenig, Clément, & Harris, 2004). By the end of their first year of life, infants selectively point to objects when informants prove themselves knowledgeable (Grassmann & Tomasello, 2010) and by 24 months old, toddlers are sensitive to informant differences across a wide variety of linguistic and non-linguistic cues like confidence (Juteau, Cossette, Millette, & Brosseau-Liard, 2019), knowledge (Luchkina, Sobel, & Morgan, 2018), and native language (Begus, Gliga, & Southgate, 2016).

For example, Koenig and Woodward (2010) exposed 24-month-olds to speakers who either provided a correct or incorrect label for a highly familiar object. Both speakers then taught novel labels for novel objects. Toddlers endorsed the label from the reliable speaker, and dispreferred the label taught by the unreliable speaker. Subsequent work also demonstrates that toddlers retrospectively reassess their endorsement of a speaker's label if that speaker later proves unreliable (Dautriche, Goupil, Smith, & Rabagliati, 2021), and preferentially generalize labels that were learned from reliable speakers to unfamiliar speakers (Luchkina et al., 2018).

However, these findings do not show that toddlers disregard unreliable speakers completely. If tested immediately after exposure to the novel labels, toddlers retain labels from unreliable and reliable speakers, but after a short delay toddlers more reliably retain the novels taught by reliable speakers (Gangopadhyay & Kaushanskaya, 2022; Mangardich & Sabbagh, 2018; Sabbagh & Shafman, 2009). Together, these lines of work suggest that reliability effects may be strongest when toddlers are forced to endorse one speaker over the other in a forced choice paradigm, as typically used in the preferential learning literature. Additionally, by using higher-level tasks such as word learning as the primary outcome measure, these studies rely on additional skills that are still developing at this stage of language acquisition. Finally, it remains unclear how toddlers' assessments of speaker reliability affects related cognitive processes such as social information-seeking.

The current study leverages toddlers' information-seeking behavior to measure their preferences for sampling novel labels from reliable and unreliable speakers. Recent advances in eye-tracking technology have enabled the development of a method that allows infants and toddlers to elicit informa-





Image Stimuli:				
Reliable Label:	<b>ball</b>	<b>book</b>	<b>duck</b>	<b>apple</b>
Unreliable Label:	<b>hat</b>	<b>cat</b>	<b>train</b>	<b>dog</b>

Figure 1: Four pairs of highly familiar objects were selected for the familiarization phase. The object pairs and their corresponding unreliable labels were selected to balance their saliency, animacy, object category, and average age of acquisition. All object label pairs and their uninformative counterparts were selected to maximize visual and phonetic discriminability.

tion from images on a screen using their eye-gaze alone. Gaze-contingent eye-tracking (Wang et al., 2012; Deligianni, Senju, Gergely, & Csibra, 2011) has been used to show that toddlers are more likely to seek the labels of objects for which they have less information (Zettersten, 2020). We suggest that gaze-contingent active sampling may also be useful for measuring finer-grained information-seeking preferences beyond the word-knowledge domain. In this case, we are interested in using gaze-contingent sampling preferences to measure toddlers' understanding of *informativity*.

To do so, we first familiarized toddlers to informants who always provide either reliable or unreliable labels for highly familiar objects. We then allow toddlers to sample novel labels for novel objects from these informants in a gaze-contingent eye-tracking task. The dynamic nature of this task allows us to explore both the toddlers' first sample preference and their overall rate of sampling throughout the entire sampling phase. These simple eye-gaze measures provide insight into how complex social inferences such as speaker informativity may affect lower-level, automatic behaviors such as information-seeking.

## Methods

### Participants

Eighteen toddlers participated in the study to date ( $M = 23mo$ , age range: 22.2 – 24.2mo; data collection is ongoing with a planned N of 40). Participants were recruited from a database of families in an urban area of the Midwestern United States. Exclusion criteria include hearing loss, vision impairments, known developmental disabilities, and/or more than 10h/week of non-English language exposure. An additional 13 participants were excluded from the analysis due to fussiness ( $N = 4$ ) or issues during eye-tracker calibration ( $N = 9$ ).

### Stimuli

The familiarization stimuli include four pairs of familiar objects, their corresponding labels, and four additional familiar

labels. The four familiar object pairs are: *ball* and *car*, *truck* and *apple*, *bear* and *duck*, *shoe* and *book* (see Figure 1). The additional familiar words used as incorrect labels are *dog*, *hat*, *train*, and *cat*. These items were chosen to be familiar (at least 50% producing the word) to 22-month-old infants according to Wordbank production trajectories (Frank, Braginsky, Yurovsky, & Marchman, 2017). The objects were paired to maximize phonetic and visual differentiability while balancing animacy, visual interest, and average age of acquisition. Test stimuli include two novel objects and four labels (*coodle*, *minu*, *toma*, and *modi*) selected from the NOUN Database (Horst & Hout, 2016).

### Parental Questionnaires

We administered two parental questionnaires. First, we used the MacArthur-Bates Communicative Development Inventory - Short Form (Level II, Form B) (Fenson et al., 1994) to measure overall vocabulary size. To ensure the familiar objects were sufficiently familiar to the participants, we asked parents to complete a survey about their child's understanding and production of the specific items used in the study.

### Procedure

Participants were tested in the laboratory while seated on a caregiver's lap. The caregiver wore darkened glasses to minimize bias and interference in the study. Throughout the study, the toddlers' eye movements were recorded with a Tobii T60 XL eye tracker. The eye tracker is first calibrated to each toddler before the study, after which the full task is administered in three phases: (1) gaze-contingent training phase, (2) speaker exposure phase, and (3) gaze-contingent sampling phase.

**Gaze-contingent Training Phase** Because this procedure requires toddlers to actively control the screen in a manner they may not be used to, we begin the study with a short gaze-contingent training phase. During this phase, two images of scenes appear on the left and right side of the screen in full color for 1 second. They then turn grayscale and the

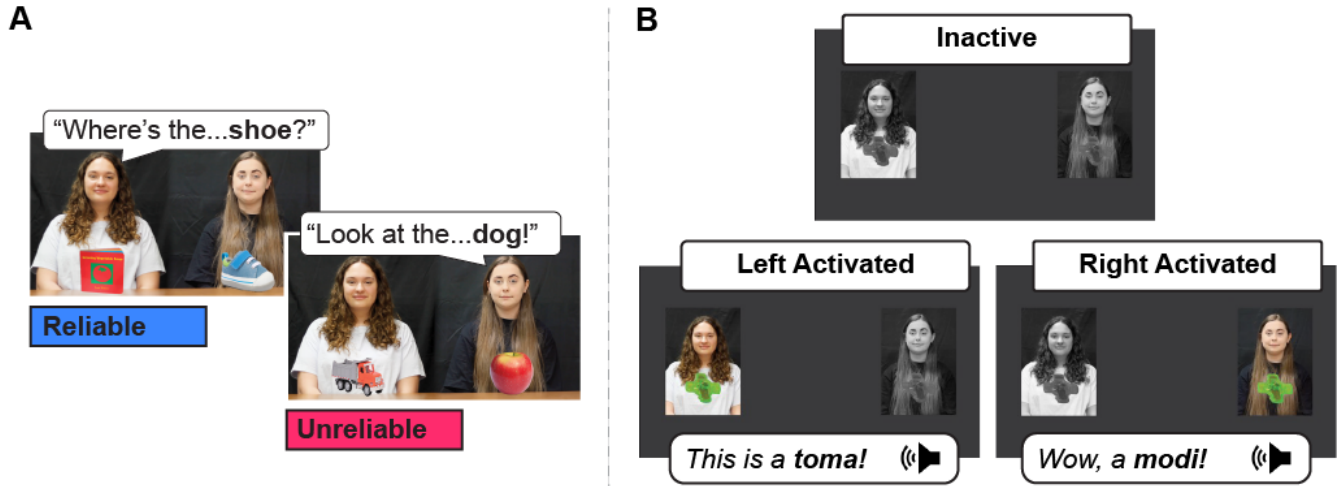


Figure 2: (A) Example trials from the speaker familiarization phase. The reliable speaker always provides a relevant label and the unreliable speaker always provides an irrelevant label. The side of the target object is sudo-randomized across trials. (B) Example gaze-contingent sampling trial. As participants fixate on the images of the two speakers, the image lights up and the speaker's voice begins to label the novel object.

gaze-contingent phase begins. At this point, when the toddler fixates on either of the two images for at least 333ms, infant-friendly music begins to play. To discourage lingering on only one image for the entire session, if the infant is still fixated on the image after 10 seconds, the audio ends and the image returns to grayscale. The infant can then direct their attention to the other image or re-trigger the first image. The stimuli respond in this gaze-contingent manner for 20 seconds. This phase ensures after which a 5-second attention-grabbing video plays and the study continues to the speaker familiarization phase.

**Speaker Familiarization Phase** In this phase, participants are exposed to two speakers who differ in the reliability of the labels they provide. The videos feature two females seated at a table with a pair of familiar objects centered in front of them. The speakers are the same age and race but wear two different colored shirts to help toddlers visually differentiate between them. The speakers took turns providing a label. Critically, one speaker always provides a reliable label (e.g., "Where's the ... shoe?" in the presence of a book and a shoe) while the other speaker provides an unreliable label (e.g., "Find the ... dog!" in the presence of a truck and an apple). Neither speaker provides any additional referential information such as eye gaze or pointing to disambiguate between the objects. Figure 2A demonstrates two example trials where the reliable speaker is seated on the left and the unreliable speaker is seated on the right.

All four pairs of objects are labeled a total of four times - twice by the reliable speaker and twice by the unreliable speaker. Thus, all four pairs of objects received both a reliable label and an unreliable label. This design was intended

to heighten any potential effect of speaker reliability as the toddler should be able to compare the behavior of the two speakers given conflicting testimony (Vanderbilt, Heyman, & Liu, 2014). To further maximize their ability to evaluate the unreliable speaker, we ensured that each familiarization phase began with a reliable speaker trial. This allowed toddlers to evaluate the unreliable speaker in the context of a reliable example (Gweon & Asaba, 2018).

The order of appearance for each pair of objects was pseudo-randomized across the experiment so that the same object pairs never appeared twice in a row and trials with the same speaker type (reliable vs unreliable) never appeared more than twice in a row. Four carrier phrases (i.e., "Do you see the...", "Can you find the...", "Where's the...", and "Look at the...") were sudo-randomized and counterbalanced per speaker type. The position of the experimenter (left vs right) and their reliability status (reliable vs unreliable) were counterbalanced across participants to control for speaker preference or side bias. The side of the labeled object is also pseudo-randomized across trials. On unreliable trials where the target object is undefined or ambiguous, a target side is randomly assigned for analysis purposes.

**Contingent Sampling Phase** The toddlers then participate in a gaze-contingent task where they can sample information about a novel object from either of the two speakers. On each of two trials, still images of the two speakers appear on the screen with a novel object superimposed on each image. The same novel object is shown with both speakers to further encourage preferential information seeking. After one second of full color, the images turn to grayscale and the gaze-contingent phase begins. As in the gaze-contingent

training phase, the toddlers have the opportunity to use their gaze to trigger information from either of the two images on the screen.

When a toddler fixates on either image for longer than 333ms, that speaker's image turns to full color and audio of the speaker labeling the novel object begins to play (e.g., "Look, a minu! I like the minu!"). This audio plays twice or until the toddler looks away, whichever comes first, at which point the audio ends and the image returns to grayscale. If the toddler then fixates on the other speaker's image, that speaker's voice begins to play with a different label (e.g., "Wow, a toma! Do you like the toma?"). The speakers provide two conflicting labels so there is contrasting information from each speaker on each trial. Each trial lasts 20 seconds with a short (5s) attention-getting video between the trials. See Figure 2B for a schematic of the gaze-contingent phase.

**Exclusion Criteria** Familiarization trials with at least 50% track loss were excluded from the analyses. After applying this criterion, only 12 of 288 (4.17%) familiarization trials were excluded. On average, each participant contributed data from 7.7 familiarization trials featuring a reliable speaker and 7.55 trials featuring an unreliable speaker. Additionally, gaze-contingent trials were only included in these analyses if at least 70% of the eye-tracking data for the full 20-second trial is usable. The responsive nature of the stimuli during the gaze-contingent phase requires high-quality eye-tracking data; this stringent criterion ensures that all participants had an unambiguously contingent experience during the sampling phase. After applying this criterion, 3 of 36 gaze-contingent trials were excluded from the analyses. On average, participants contributed 18.79 sec out of a possible 20 sec (91.72%) of gaze-contingent data for each trial.

## Results

Because our data only represent a portion of the total pre-registered sample of  $N = 40$ , only summary statistics and descriptive analyses have been reported here. Preregistration, analysis code, and experiment code can be found at: <https://osf.io/4nw57/>.

### Children accurately track the reliability of speakers.

We begin by considering a basic manipulation check. In this study, participants must be able to track the informativity of the two speakers during the familiarization phase. To form an association between a speaker and their demonstrated reliability, participants must attend to both the speaker and the presence (or absence) of the labeled object. If a participant is only attending to speaker faces but not objects (or objects, but not speakers), they should not be able to track an individual's label reliability. To check that participants are sufficiently attending to all key areas of the familiarization phase, gaze coordinates from each eye were averaged and mapped onto four areas of interest (AOIs): the two faces of the speakers and the two objects. The proportion of looks to each of the four AOIs was calculated after removing frames of missing gaze data or

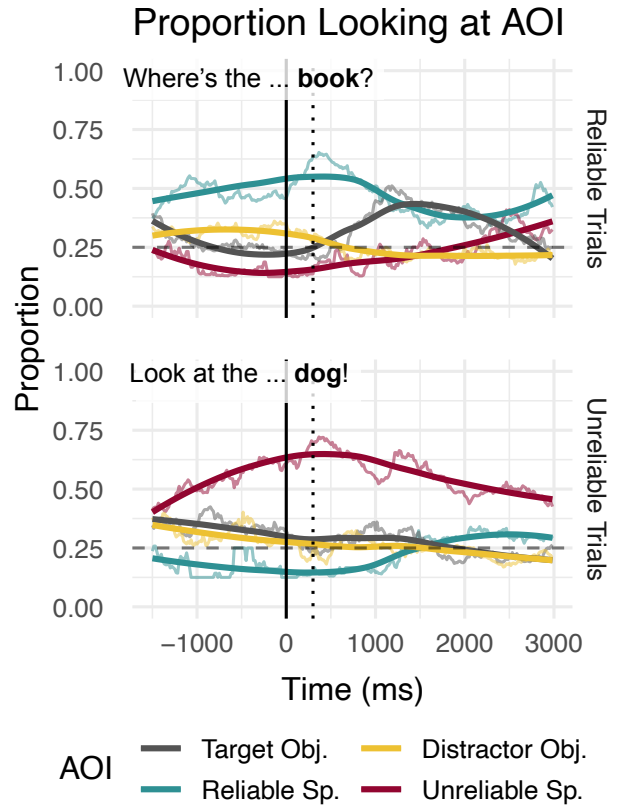


Figure 3: Proportion looking towards each area of interest (AOI) during the 16 familiarization trials. Trials are averaged by participants and averaged per 16.66ms timebin. Chance proportion looking is denoted by the dashed horizontal line at .25. The solid vertical line signifies the label onset at 0ms and the dotted vertical line marks 300ms after the onset of the target word and the beginning of the critical window.

looks to other portions of the screen.

Figure 3 shows the proportion of gaze towards each AOI for each 16.66ms time bin during the familiarization phase. If participants are attending to all four AOIs equally, their proportion looking would be at chance, indicated by the dashed line at .25. Each trial features a speaker labeling an object with a carrier phrase (e.g., "Find the...[label]!"). The solid line at 0ms represents the onset of the beginning of the critical window. As expected, participants are attending to the relevant speaker during the labeling events - they fixate on the reliable speaker on reliable trials and the unreliable speaker during unreliable trials.

We also see an increase in looks toward the target object only on reliable trials. This confirms that participants are locating the relevant target objects during reliable labeling events and are unable to identify a target object during unreliable labeling events. After restricting the data to the key

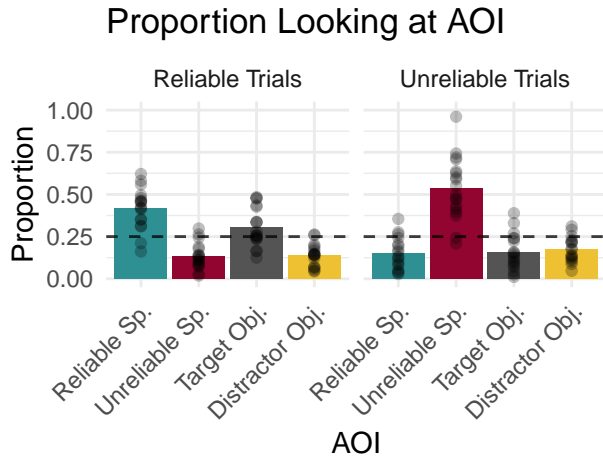


Figure 4: Mean proportion of looks to each area of interest (AOI) averaged across the critical window (300ms - 3000ms). Data points represent the average proportion of looks for each participant. Chance-looking behavior (looking equally to all four AOIs) is represented by the horizontal dashed line at .25.

analysis window after the label onset (Fig. 4), we find that participants tend to look to the target object about twice as often during reliable speaker trials ( $M = .306$ , 95% CI = [0.25, 0.36]) compared to unreliable speaker trials ( $M = .15$ , 95% CI = [0.10, 0.20]).

### Children look to the reliable speaker first.

Because participants can only trigger information from one speaker at a time, their sampling choice in the first few moments of each gaze-contingent trial should represent their information-seeking priorities. If they are using speaker reliability to prioritize their sampling, we should expect them to privilege sampling information from the reliable speaker over the unreliable speaker. To quantify a first-look preference, we first filtered out fixations shorter than 300ms to remove any looks generated by shifting gaze. We then identified the onset of the first fixation toward either of the two speakers. As shown in Figure 5, participants first fixated on the informative speaker at 1315.65ms (95% CI [924.15, 1707.16]) while they first fixated on the uninformative speaker at 1502.02ms (95% CI [1037.32, 1966.71]). These results suggest that toddlers may use speaker reliability to privilege their social information-seeking preferences.

### Children sample evenly from each speaker.

Because the gaze-contingent sampling phase provides ample time to sample from either speaker repeatedly, we are also interested in toddlers' overall sampling preferences. If toddlers are using speaker reliability to filter and restrict their attention to only reliable speakers, we should see an effect on

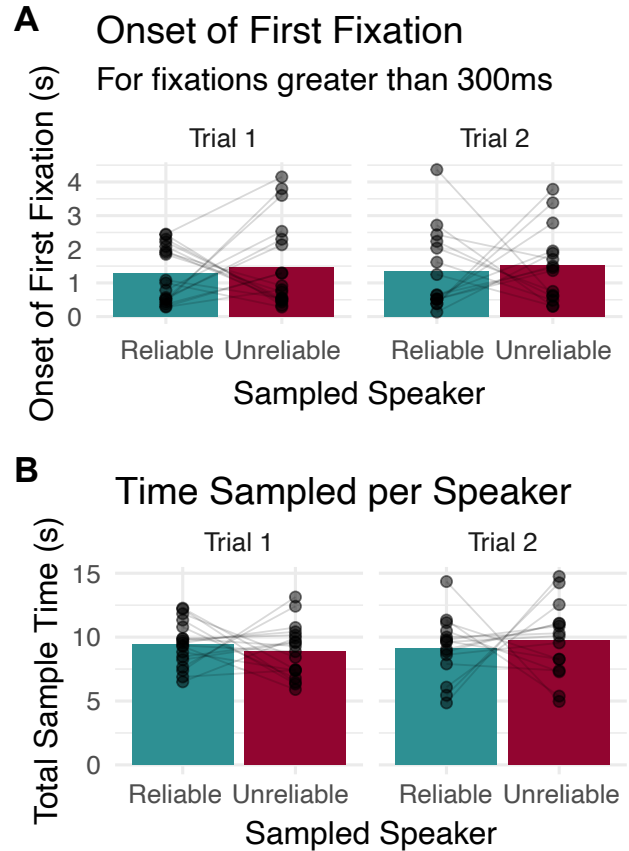


Figure 5: (A) The average onset of the first fixation towards either of the two speakers (reliable in blue, unreliable in red) split by trial number. Each point represents a single participant. (B) The average total sampling time for each speaker type, separated by trial number. Each bar represents the mean sampling time for each trial, for each speaker type (reliable vs unreliable). Each point represents a single participant.

their sampling behavior. Alternatively, if information from unreliable speakers is still informative or still worth exploring, toddlers may still sample from both speakers throughout the sampling phase. To test these predictions, we averaged the gaze coordinates from each eye and assigned two areas of interest: the two images of the speakers holding novel objects. Figure 5B shows the average time participants spent sampling from either of the two speakers for each sampling trial. These data suggest there is no significant sampling preference for reliable speakers ( $M = 9.41s$ , 95% CI [8.98, 9.85]) or unreliable speakers ( $M = 9.11s$ , 95% CI [8.55, 9.66]). In this case, the information provided by the uninformative speaker may have been equally interesting to the toddlers when the pressure of time or a forced choice is lifted.



## Discussion

Toddlers and young children have strong socio-pragmatic abilities which allow them to assess the quality or quantity of information they can receive from people in their environment (Harris et al., 2018; Grassmann & Tomasello, 2010; Luchkina et al., 2018; Begus et al., 2016). They are particularly sensitive to the reliability of speakers in their environment, and prefer to learn novel labels from reliable speakers over unreliable speakers (Koenig & Woodward, 2010; Koenig et al., 2004; Dautriche et al., 2021). However, young children still retain novel labels from unreliable speakers immediately after training (Gangopadhyay & Kaushanskaya, 2022); only after a delay is the reliable speaker's label retained and lexicalized and the unreliable speaker's label dropped (Sabbagh & Shafman, 2009).

It remains unclear how speaker reliability guides toddlers' real-time information-seeking behavior. With a novel application of a gaze-contingent active sampling paradigm, we found evidence that 23-month-old toddlers may privilege novel object label information from reliable speakers over unreliable speakers. In the first moments of the gaze-contingent task, toddlers are more likely to defer first to the reliable speaker followed by the unreliable speaker. However, when given ample opportunity to sample from either speaker over a 20-second window, toddlers were no more attentive to reliable speakers over unreliable speakers. Our results suggest that young children do not use speaker reliability to restrict their information-seeking preferences. Importantly, these data represent just over 40% of our planned sample.

If this pattern of data holds up with a larger sample, there are several reasons why early language learners may continue to monitor information from unreliable sources. For one, speakers who are known to be unreliable may still provide information that is useful for language learners. Toddlers may be sampling from the unreliable speaker as a form of counterfactual information seeking (Fitzgibbon & Murayama, 2022). While the speaker may not provide information about the *correct* label that belongs to a novel object, they are still providing information about what is likely to be an *incorrect* label for a novel object.

Alternatively, it could still be useful to learn the unreliable speaker's label for future communication purposes. While early language experience is typically characterized as having a single label for each object, in natural language environments word labels and referents are highly context and speaker-dependent. Even very young children can associate specific words with specific speakers and expect interlocutors to be referentially consistent (Graham, Sedivy, & Khu, 2014). Furthermore, objects may be referred to in different ways depending on the situating context (Wojcik, Zettersten, & Benitez, 2022; Callanan & Sabbagh, 2004). Because language is flexible and context-dependent, sampling more information may always be preferable, regardless of the information quality from the informant.

**Limitations and Future Directions** Our experiment differs from prior work in a few significant ways that limit the interpretability of our findings. First, the speakers in this task provided no supporting information about the intended referent on each trial (e.g., pointing or eye gaze). This means the toddler must not only attend to both possible referents on the screen, but also assess the appropriateness of the label provided by the speaker. This difference is most important during the unreliable speaker trials. While the unreliable speaker is providing less informative labels compared to the reliable speaker, they are not unambiguously labeling the objects in the scene incorrectly (i.e., they are not explicitly pointing at an apple and calling it a "dog", they are just saying the word "dog"). This ambiguity may make it more difficult for toddlers to assess the relative reliability of the speakers. However, by removing the additional non-linguistic referential cues, we can ensure that any speaker preferences are formed based on the linguistic content alone.

The ambiguity in the unreliable trials also leaves open the possibility that the speaker is labeling objects outside of the context or off-screen. If the toddlers assume that both speakers are felicitous, they may infer that the unreliable speaker is providing additional information that is not available to the participant. However, the toddlers in this experiment do not seem to search outside of the context for additional possible referents - instead, they shift their gaze between the two possible references in context.

Second, because the scope of this study was strictly focused on toddlers' information-seeking preferences, there was no additional measure of which word (if any) the toddlers are more likely to retain. It may be the case that, despite their equal rates of sampling, they would use the sampled information from the two speakers differently in a downstream word-learning task. That is, toddlers could be sampling and attending to the uninformative speaker, but would preferentially learn and endorse the label provided by the reliable speaker. Future work could address this question by including a preferential word learning test after the gaze-contingent sampling phase to measure their word learning preferences.

This paradigm provides a promising approach to studying multiple dimensions of information-seeking and social learning. In future work, we can manipulate other aspects of informativity beyond reliability such as under- and over-informativity. This method also enables us to study how other speaker characteristics such as expertise or familiarity may influence toddlers' information-seeking and decision-making processes.

## Acknowledgments

We would like to thank all of the participating families, the Infant Learning Lab research assistants, and our generous funders. This work was funded by grants from the National Institute of Child Health and Human Development to JRS (R37HD037466) and the Waisman Center (U54 HD090256).

## References

- Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences*, 113(44), 12397–12402. doi: 10.1073/pnas.1603261113
- Begus, K., & Southgate, V. (2018). Curious learners: How infants' motivation to learn shapes and is shaped by infants' interactions with the social world. *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms*, 13–37.
- Callanan, M. A., & Sabbagh, M. A. (2004). Multiple Labels for Objects in Conversations With Young Children: Parents' Language and Children's Developing Expectations About Word Meanings. *Developmental Psychology*, 40(5), 746–763. doi: 10.1037/0012-1649.40.5.746
- Chu, J., & Schulz, L. E. (2020). Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, 2, 317–343.
- Dautriche, I., Goupil, L., Smith, K., & Rabagliati, H. (2021). Knowing How You Know: Toddlers Reevaluate Words Learned From an Unreliable Speaker. *Open Mind*, 5, 1–19. doi: 10.1162/opmi.a.00038
- Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental psychology*, 47(6), 1499.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), i. doi: 10.2307/1166093
- Fitzgibbon, L., & Murayama, K. (2022). Counterfactual curiosity: Motivated thinking about what might have been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1866), 20210340. doi: 10.1098/rstb.2021.0340
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. doi: 10.1017/S0305000916000209
- Gangopadhyay, I., & Kaushanskaya, M. (2022). The effect of speaker reliability on word learning in monolingual and bilingual children. *Cognitive Development*, 64, 101252. doi: 10.1016/j.cogdev.2022.101252
- Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, 17(11), 585–593.
- Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, 41(1), 34–50. doi: 10.1017/S0305000912000530
- Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference: Primacy of pointing. *Developmental Science*, 13(1), 252–263. doi: 10.1111/j.1467-7687.2009.00871.x
- Gweon, H., & Asaba, M. (2018). Order Matters: Children's Evaluation of Underinformative Teachers Depends on Context. *Child Development*, 89(3). doi: 10.1111/cdev.12825
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, 69, 251–273.
- Hembacher, E., deMayo, B., & Frank, M. C. (2020). Children's social information seeking is sensitive to referential ambiguity. *Child Development*, 91(6). doi: 10.1111/cdev.13427
- Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (noun) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. doi: 10.3758/s13428-015-0647-3
- Juteau, A.-L., Cossette, I., Millette, M.-P., & Brosseau-Liard, P. (2019). Individual Differences in Children's Preference to Learn From a Confident Informant. *Frontiers in Psychology*, 10, 2006. doi: 10.3389/fpsyg.2019.02006
- Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, 88(3), 449–460.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological science*, 15(10), 694–698.
- Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, 46(4), 815–826. doi: 10.1037/a0019664
- Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, 201, 104315.
- Lee, C., & Lew-Williams, C. (2023). The dynamic functions of social cues during children's word learning. *Infant and Child Development*, 32(1), e2372.
- Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science*, 21(6). doi: 10.1111/desc.12663
- Mangardich, H., & Sabbagh, M. A. (2018). Children remember words from ignorant speakers but do not attach meaning: Evidence from event-related potentials. *Developmental Science*, 21(2), e12544. doi: 10.1111/desc.12544
- Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child development*, 105–114.
- Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers. *Cognition*, 112(3), 415–422. doi: 10.1016/j.cognition.2009.06.005
- Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, 18(9), 668–671. doi: 10.1016/j.cub.2008.03.059
- Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science*, 17(3), 443–

451. doi: 10.1111/desc.12134

- Waismeyer, A., & Meltzoff, A. N. (2017). Learning to make things happen: Infants' observational learning of social and physical causal events. *Journal of experimental child psychology*, 162, 58–71.
- Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PloS one*, 7(2), e30884.
- Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *WIREs Cognitive Science*, 13(4). doi: 10.1002/wcs.1596
- Zettersten, M. (2020). *Seeking new words: Active word learning in infants and children*. PhD Diss., University of Wisconsin-Madison.