Journal of Experimental Child Psychology 189 (2020) 104701

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Brief Report

I don't believe what you said before: Preschoolers retrospectively discount information from inaccurate speakers

Elena Luchkina [a,*], Kathleen H. Corriveau [b], David M. Sobel [c]

[a] University of California, Berkeley, Berkeley, CA 94720, USA
[b] Boston University, Boston, MA 02215, USA
[c] Brown University, Providence, RI 02912, USA

ARTICLE INFO

Article history:
Received 9 November 2018
Revised 5 August 2019
15 August 2019
Available online 8 October 2019

Keywords:
Selective word learning
Speaker reliability
Retrospective reliability inference
Preschoolers' vocabulary development
Language development
Social reasoning

ABSTRACT

Children use speakers' past accuracy to make inferences about novel word meanings those individuals provide in the future. An open question is whether children can retrospectively reevaluate information after learning that the source was inaccurate. We addressed this question in two experiments where a speaker first introduced labels for novel objects and then revealed that she is either accurate or inaccurate in naming familiar objects. Experiment 1 showed that 3.5- to 6.5-year-olds displayed enhanced performance on a word knowledge test when they had learned novel words from a speaker who then showed herself to be an accurate labeler as opposed to an inaccurate labeler. Experiment 2 replicated this procedure but had a different speaker provide inaccurate label information. This manipulation did not affect learning, suggesting that children discount speakers and are not simply influenced by the demands of processing inaccurate information. Together, these results indicate that 3.5- to 6.5-year-olds continue to monitor the speakers' accuracy after learning new words from them, update their beliefs as accuracy data become available, and selectively retain words learned from speakers who they deem to be epistemically competent.

* Corresponding author.
E-mail address: elenaluchkina@berkeley.edu (E. Luchkina).

https://doi.org/10.1016/j.jecp.2019.104701

Introduction

Numerous studies have demonstrated that preschoolers learn the meaning of novel words selectively based on speakers' past accuracy (e.g., Clément, Koenig, & Harris, 2004; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; see Harris, 2012, for a review). Toddlers have selective learning capacities as well, both for the meaning of novel words (e.g., Brooker & Poulin-Dubois, 2013; Luchkina, Sobel, & Morgan, 2018; Koenig & Woodward, 2010) and for nonverbal information (e.g., Over & Carpenter, 2012). By 4 years of age, children not only selectively learn words from accurate speakers over inaccurate speakers but also take various social factors into consideration when making inferences about credibility (e.g., Kushnir, Wellman, & Gelman, 2008; Nurmsoo & Robinson, 2009; see Sobel & Kushnir, 2013, for a review).

These conclusions about children's selective learning capabilities center on procedures in which speakers' accuracy is established prior to the introduction of novel object labels. In most word learning situations, however, children rarely know speakers' reliability ahead of time. The question we ask here is whether children can discount novel linguistic information retrospectively after learning that the source of knowledge is subsequently unreliable.

One previous study investigated this question. Scofield and Behrend (2008, Experiment 2) showed that preschoolers could revise their judgments about novel word meanings after observing that the speaker who introduced them was subsequently inaccurate in naming familiar objects. However, only 10 of the 24 tested children (42%) in this study demonstrated this behavior. In a separate experiment (Experiment 3) where children were retested after a delay, 22 of the 42 4-year-olds tested (52%) revised their judgments about the label of a novel object based on subsequent information about the reliability of the informant. These results suggest that children potentially revise their judgments about word labels based on information about a speaker's reliability, but these abilities might be nascent during the preschool years or available to only a subset of the sample.

Note also that the children in Scofield and Behrend (2008) study were presented with two informants. Both informants generated different novel labels for the novel object. One informant then accurately labeled familiar objects, whereas the other inaccurately labeled those objects. An open question is whether children infer the reliability of each informant or whether they responded based on the contrast between these informants' information. Some research indicates that when children are presented with only a single informant, they do not show a selective preference for learning novel words from a previously accurate informant over a previously inaccurate informant (Vanderbilt, Heyman, & Liu, 2014). Thus, without the conflict between accurate and inaccurate labeling, children might not revise their inferences about novel labels based on the subsequent accuracy of informants.

In the current study, we built on Scofield and Behrend (2008) findings using a single-informant paradigm. This allowed us to examine whether children could reevaluate the novel label associated with a novel object after subsequently receiving information about the reliability of a speaker's knowledge. Our rationale for choosing a single-informant design was to present children with a situation that provided minimal cues to potential informant unreliability. When two informants provide conflicting labels for the same object, children may have hypothesized that at least one of those informants was unreliable. Thus, they should pay close attention to the informant's subsequent verbal behaviors. Our interest was in evaluating whether children spontaneously engage in accuracy-monitoring behavior and apply inferences from speakers' accuracy in labeling to the previously taught information. Such a scenario allowed us to establish a more stringent test of children's active monitoring of the quality of their sources of information—that is, whether children keep track of informants' reliability when they have no particular reason to doubt it. This procedure was also a more realistic measure of how children might learn novel words; rarely do children encounter different informants providing different novel labels for the same object in quick succession.

In Experiment 1, we explored children's capacity to remember the referents of novel labels generated by informants who then subsequently revealed themselves to be accurate or inaccurate speakers. We also asked children to make inferences about the labels for other objects using a disambiguation paradigm (e.g., Markman & Wachtel, 1988). Children use the reliability of a speaker to make such linguistic inferences (Diesendruck, Carmel, & Markson, 2010) and relate their inference to their retention of the label–object mapping (Sobel, Sedivy, Buchanan, & Hennessy, 2012).

We also considered a potential mechanism for children's retrospective capacities. Some researchers have argued that children encode but do not generalize novel information provided by unreliable speakers (e.g., Henderson, Graham, & Schell, 2015) or reject such information right away (e.g., Mangardich & Sabbagh, 2018; Sabbagh & Shafman, 2009; Sabbagh, Wdowiak, & Ottaway, 2003). On this interpretation, information is encoded but then discounted after learning that the source is unreliable. Other researchers suggest that the increased demands of processing incorrect information affect children's ability to retain novel information provided by unreliable speakers (e.g., Barry, 2016; Bernard, Proust, & Clément, 2014; Doebel, Rowell, & Koenig, 2016). This latter view suggests that hearing inaccurate labels for familiar objects potentially interferes with the processing of novel lexical information. This possibility could also be extended to retrospective inferences by considering the interference in a retrospective manner (when speakers' accuracy is revealed after children learned new information from the speakers).

We consider these alternatives in Experiment 2 by replicating our procedure from Experiment 1 but substituting a novel speaker who provided accurate or inaccurate labels for familiar objects in the place of the informant who labeled the novel objects. If information is encoded but discounted through establishing speaker reliability, then children in Experiment 2 should retain the novel labels. If processing incorrect information interferes with encoding of novel labels, then children should be unable to retain them.

Experiment 1

Experiment 1 extended a procedure developed by Scofield and Behrend (2008) to a group of 3.5- to 6.5-year-olds. Scofield and Behrend found that about half of the 4-year-olds they tested made retrospective inferences about speakers' reliability. In the current study, we included children of a broader age range to ensure the robustness of this capacity. Moreover, instead of contrasting information from accurate and inaccurate informants, we presented children with a single informant who initially generated labels for novel objects and then accurately or inaccurately labeled familiar objects. This ensured that children were reevaluating their belief about object labels based on subsequent information about an informant's reliability rather than responding based on the conflict between informants. We also included two different measures of children's learning: their ability to retain the novel labels provided by the speakers and their ability to make an inference about the labels of different novel objects based on the principle of mutual exclusivity.

Methods

Participants

A total of 48 monolingual English-speaking children (24 per condition) aged 3.5–6.5 years participated in the study (M_{age} = 60 months, range = 42–81; 19 girls and 29 boys). Although ethnicity data were not collected, the majority of children were Caucasian and represented middle- or high-SES (socioeconomic status) backgrounds. Children were recruited and tested at a local children's museum and at the university laboratory. One additional child was tested but excluded from analyses due to the loss of interest in the task.

Apparatus

Novel label training and accuracy demonstration were video-recorded and presented on a MacBook 11-inch laptop computer screen to ensure minimal differences in children's exposure to the stimuli other than the experimental manipulation. Auditory stimuli were played through the laptop speakers set to the maximum volume. Test trials were conducted live by an experimenter who sat across the table from the children.

Materials

Six novel objects were used as test stimuli: a wooden hair pick, a folded vinyl soap holder, a white plastic pipe hook, a mental plug wrench, a yellow tarpaulin clip, and a green plastic drain cover (see Fig. 1). The green drain cover and metal plug wrench were the novel objects presented and labeled during novel label training. The remaining objects were first introduced during test trials. Three familiar objects were used during accuracy demonstration: a sneaker, a golden cardboard star, and a stuffed dog.

Procedure

The experimental procedure was designed and conducted in accordance with the approved protocol issued by the institutional review board at Brown University. Children were tested in a quiet room located at the museum or at the laboratory. The experiment included three phases presented in a fixed order—*novel label training, accuracy demonstration,* and *test*—that are described in more detail below (see Fig. 2).

In novel label training, children were first seated across the table in front of an experimenter who explained the task to them and started the video. While children watched the video, the experimenter remained in the back of the room so that the children could not see her. In each trial, a female actor on the video introduced two novel objects and labeled them "lif" and "wug" (e.g., "Look, a wug! Wug!"). Each object was labeled twice and manipulated twice, so that each label–object pair was heard four times total. Label–object pairings and the order of object presentation were counterbalanced among participants.

Following novel label training, children were assigned to one of two conditions for the accuracy demonstration trials. Children in the *accurate speaker* condition saw the same actor from the novel label training phase introduce three familiar objects and label them accurately. Children in the *inaccurate speaker* condition saw the actor label the same objects inaccurately with familiar inaccurate labels (e.g., "cat," "book," "cup"). Each familiar label was mentioned twice (e.g., "Look, a cat! Cat!"). During each trial, the actor alternated between looking at the object and looking at the camera to emulate eye contact and establish joint attention.

Immediately following the accuracy demonstration, the experimenter removed the laptop and sat across from the children to administer four test trials—two *retention* trials (one per novel label) and two *disambiguation* trials—presented in alternation. On each test trial, three objects—the two novel objects from the video and one unfamiliar novel object—were placed in front of the children. The unfamiliar novel object was different on every trial. On retention trials, the experimenter invited children to choose the referent associated with the label used during the novel label training trials (e.g., "Which one is the lif/wug?"). On disambiguation trials, the experimenter invited children to choose the referent associated with an unfamiliar novel label (e.g., "Which one is the dax?"; "dax" and "neem" were used for these trials).

We included both retention and disambiguation trials to ensure that children's test performance did not reflect a potential tendency to select visually familiar objects. Chance-level performance with

Fig. 1. Novel objects used in novel label training and on test trials in Experiments 1 and 2.

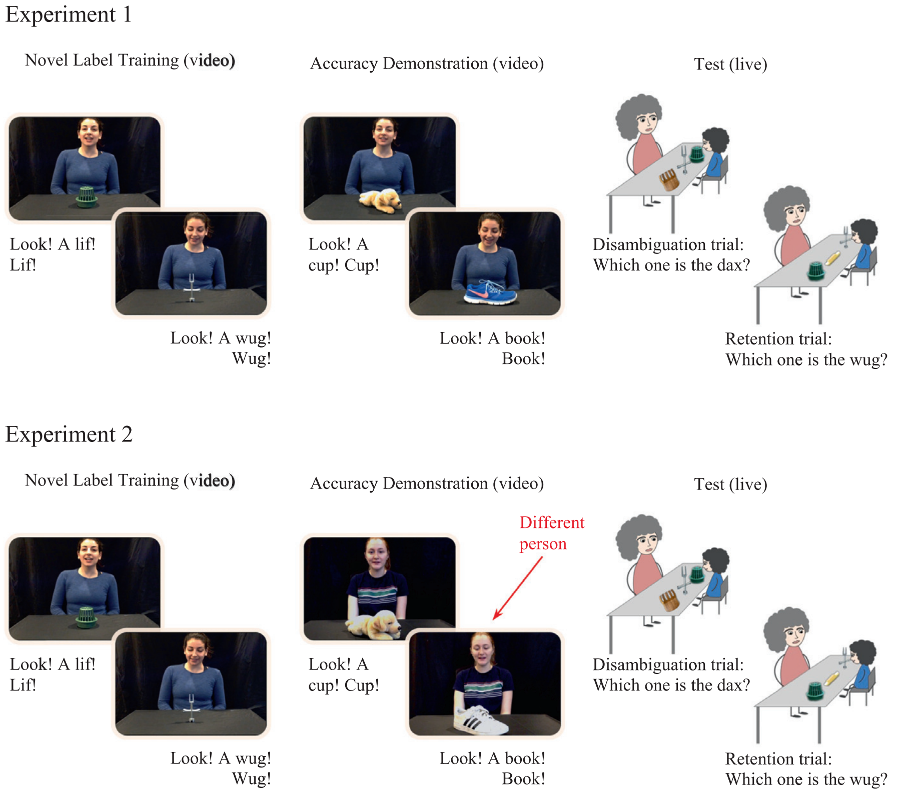

Fig. 2. Procedure in Experiments 1 and 2. Depicted here is the inaccurate speaker condition.

three possible object choices is achieved when children select the novel distractor object 33.33% of the time. If children's choices are driven by avoiding that object, then there is a possibility that even if children select the correct object only slightly more than 50% of the time, their performance can be evaluated as above chance. Such a conclusion may overestimate children's ability to retrospectively evaluate the quality of information provided to them in the past. To avoid making such an error, we included disambiguation trials on which children needed to suppress a potential impetus to select previously demonstrated objects and infer that the novel label must stand for the unnamed object. Children were not provided with any feedback on their responses other than a neutral acknowledgment ("Okay, this one"). The location of the target object was counterbalanced across trials.

Results

Responses on test trials were coded as correct (children chose the previously labeled object on retention trials or the novel object on disambiguation trials) or incorrect. We conducted preliminary analyses of correct responses by constructing a generalized estimating equation (GEE) with a binary logistic distribution to explore effects of age and gender. The overall model was not significant, nor were these factors, Wald $\chi^2(1, N = 48) = 2.74$ and Wald $\chi^2(1, N = 48) = 0.70$, $ps > .05$, respectively. These variables were excluded from further analyses.

Our main analysis focused on constructing a GEE on correct responding as the dependent measure, assuming a binary logistic distribution. Condition (accurate speaker or inaccurate speaker) was a between-participant factor. Trial type (disambiguation or retention) and trial number (first or second) were within-participant factors. This analysis revealed a significant effect of condition, Wald

$\chi^2(1, N = 48) = 7.45$, $p = .006$. Whereas children in the accurate speaker condition responded correctly on 78% of the trials, children in the inaccurate speaker condition responded correctly on only 55% of the trials. There was a significant effect of trial type, Wald $\chi^2(1, N = 48) = 4.58$, $p = .032$, with participants performing better on disambiguation trials (74% correct) than on retention trials (60%). The effect of trial number was also significant, Wald $\chi^2(1, N = 48) = 4.26$, $p = .039$, with children responding correctly on 63% of first test trials and 72% of second test trials.

Because the significant effect of trial number suggested that performance on the second retention and disambiguation trials was affected by children's inferences on the first trials, we reran all analyses only for the first trials. In this analysis, the effect of condition remained significant, Wald $\chi^2(1, N = 48) = 5.61$, $p = .018$, but the effect of trial type was no longer significant.

We also compared performance on the first retention and disambiguation trials with chance responding (33.33%; see Fig. 3). In the inaccurate speaker condition, performance on the disambiguation trial was not significantly different from chance (43%; binomial test, $p = .21$), whereas on the retention trial children chose the target objects significantly more often than chance (57%; binomial test, $p = .02$). In the accurate speaker condition, children chose the target object significantly more often than chance on both the disambiguation (68%) and retention (80%) trials (binomial tests, $ps < .001$).

Discussion

By the preschool years, children use speaker accuracy information to predict the reliability of future information provided by the same speakers. Experiment 1 suggests that they also evaluate previously learned information retrospectively. Consistent with Scofield and Behrend (2008), these results suggest that children track speakers' reliability continuously, recognize the relevance of such information to their beliefs about word meanings, and update such beliefs after receiving information about speakers' reliability.

However, to justify this conclusion, we must consider alternative explanations of performance in the inaccurate speaker condition. Instead of retrospectively discounting the information provided by the unreliable source, children might have accepted novel labels provided by accurate and inaccurate speakers equivalently but were then affected by memory demands associated with processing inaccurate labels (Koenig & Doebel, 2013); hearing incorrect labels could have interfered with children's retention of novel words.

One piece of evidence is inconsistent with this possibility, namely that we observed a difference in the types of errors children made on retention trials. Participants in the accurate speaker condition chose the distractor object 20% of the time when they made an error (4% of all responses), whereas

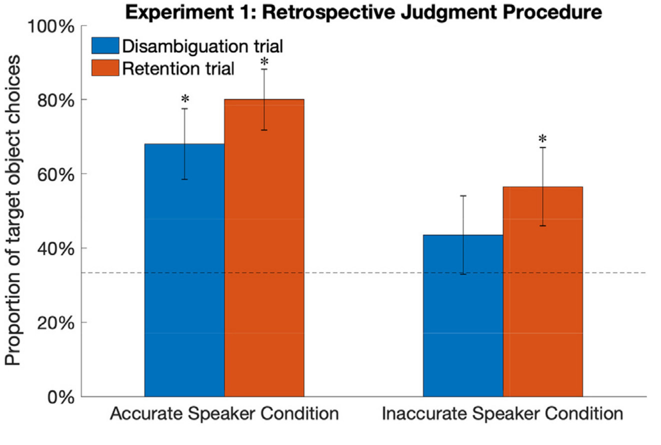

Fig. 3. Children's test performance on first trials in Experiment 1. Asterisks represent significant differences from the chance level.

the proportion of distractor choices was 70% in the inaccurate speaker condition (30% of all responses) (see Fig. 4). This difference suggests that most children thought that the inaccurate speaker had not simply mixed up the object–label pairings but instead had provided completely invalid information. Experiment 2 directly tested this possibility.

Experiment 2

Experiment 2 replicated the procedure used in Experiment 1 with one change, namely that the novel label training was presented by one speaker (Speaker 1) and the accuracy demonstration was presented by a different speaker (Speaker 2) instead of by the original speaker (as in Experiment 1). If the difference between the accurate and inaccurate conditions reflected children's retrospective inferences, then this manipulation would eliminate the effect of condition because Speaker 2's accuracy has no implications for Speaker 1's reliability. Alternatively, if the effect of condition in Experiment 1 was driven by the increased memory demands, then the effect of condition should remain because children must process the equivalent inaccurate information.

Method

Participants

A total of 50 monolingual English-speaking children (25 per condition) aged 3.5 to 6.5 years participated in the experiment (M_{age} = 62.37 months, range = 41–83; 27 girls and 23 boys). As in Experiment 1, the majority of children were Caucasian and represented middle- or high-SES backgrounds. Children were recruited and tested at a local children's museum and the laboratory.

Apparatus and materials

The apparatus and materials were the same as in Experiment 1 with the exception of the materials used in the accuracy demonstration phase. The original accuracy demonstration video clip was replaced with a clip featuring a different female actor labeling familiar objects accurately or inaccurately. The objects and labels used in the video were the same as those used in Experiment 1.

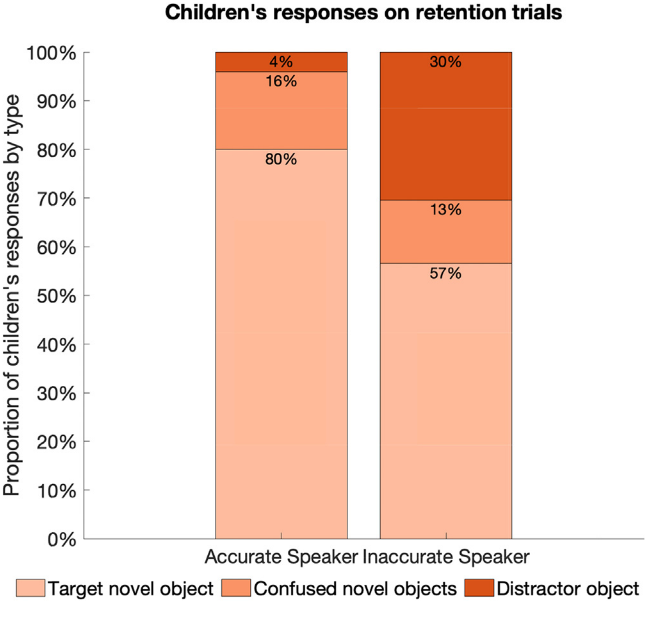

Fig. 4. Proportions of responses on the first retention trial in Experiment 1.

Procedure

The procedure was identical to the procedure in Experiment 1 (see Fig. 2) with two exceptions. First, as mentioned above, two actors were used: one in the novel object training phase and a second actor in the accuracy demonstration phase. Second, to ensure that children attended to the change of actor between the two phases, we included the following control question: "Do you remember the video you just saw? How many ladies did you see?"

Results and discussion

Of the 50 children, 8 answered the control question about the number of speakers in the video incorrectly. Their data were excluded from analyses.[1]

As in Experiment 1, preliminary analyses via GEE showed no effects of gender and age on children's choice of object, Wald $\chi^2(1, N = 42) = 0.79$ and Wald $\chi^2(1, N = 42) = 1.68$, both *p*s > .05. These variables were not considered further. For the main analysis, we constructed a GEE with a binary logistic distribution with choice of object as a dependent variable. Condition (accurate speaker or inaccurate speaker) was a between-participant factor. Trial type (disambiguation or retention) and trial number (first or second) were within-participant factors. This analysis revealed no significant effects of trial type or condition, Wald $\chi^2(1, N = 42) = 0.43$ and Wald $\chi^2(1, N = 42) = 0.04$, both *p*s > .05. As in Experiment 1, there was a significant effect of trial number, Wald $\chi^2(1, N = 42) = 3.84$, $p = .05$, with children responding correctly 63% of the time on first trials and 75% of the time on second trials.

Given this result, as in Experiment 1, we reran our analyses on the first trials only. These analyses revealed a significant effect of trial type, Wald $\chi^2(1, N = 42) = 9.09$, $p = .003$, with children responding correctly on 78% of retention trials and on 47% of disambiguation trials. There was no effect of condition on children's choice of objects during test, Wald $\chi^2(1, N = 42) = 0.60$, $p > .05$.

We also compared performance on the first retention and disambiguation trials with chance. On the retention trials, children responded differently from chance on both the accurate speaker condition (70% correct) and inaccurate speaker condition (86% correct) (binomial tests, both *p*s < .001) (see Fig. 5). On the disambiguation trials, children in both conditions did not select the target object more often than chance (children responded at 50% and 45% correct in the accurate and inaccurate conditions, respectively) (binomial tests, $p = .07$ and $p = .16$, respectively). The difference in performance on the two trial types suggests that disambiguation inferences were substantially more difficult for children than the retrieval of the novel word–object pairings.

General discussion

We investigated children's ability to make retrospective inferences about word meanings learned from speakers who subsequently revealed their accuracy or inaccuracy in naming familiar objects. Experiment 1 replicated and extended Scofield and Behrend (2008) findings using a single informant. When children learned novel labels from speakers who later mislabeled familiar objects, children were less likely to retain the object–label pairings used by those speakers on test trials than they were when they learned novel labels from speakers who named familiar objects correctly.

Two patterns of results are worth noting. First, although in Experiment 1 children in the inaccurate speaker condition performed significantly above chance on retention trials, there was a notable difference between the conditions (57% vs. 80%). It is possible that in the absence of inaccurate labeling episodes, performance on these questions was supported by both episodic memory and semantic memory, so that children remember the labeling episodes and form semantic links between novel

[1] Children who responded to the control question incorrectly performed on the word knowledge test like children in Experiment 1 and displayed a significant effect of condition, Wald $\chi^2(1, N = 8) = 69.802$, $p < .001$, and a significant effect of trial type, Wald $\chi^2(1, N = 8) = 133.854$, $p < .001$. Children in the accurate speaker condition chose the target object 100% of the time on retention trials and 80% of the time on disambiguation trials. For children in the inaccurate speaker condition, the numbers were 67% and 33% for retention and disambiguation trials, respectively. Although these analyses are on a small sample, they speak to the necessity of recognizing that the second speaker was different from the first speaker and that children were tracking informants' reliability over time.

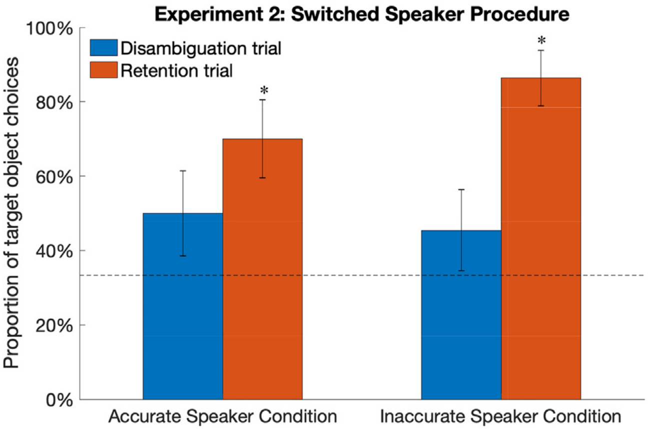

Fig. 5. Performance on first test trial in Experiment 2. Asterisks represent significant differences from the chance level.

objects and novel labels. When children witnessed inaccurate labeling following novel label training, they might engage in semantic blocking (e.g., Sabbagh & Shafman, 2009) and their above-chance performance would be driven by a residual event-based memory of a labeling event (Henderson et al., 2015).

Second, unlike in Experiment 1, performance on disambiguation trials in Experiment 2 did not deviate from chance levels in both conditions (50% and 45% for the accurate and inaccurate conditions, respectively). We hypothesize that this effect was driven by the novelty of information presented during familiarization in Experiment 2. This result also may be related to the distinction between episodic memory and semantic memory. When children were first exposed to novel labels introduced by one person and then observed a different person label familiar objects, the novelty of the social context may have interfered more with the formation of episodic memory (due to the need to store a new event) than with the formation of a novel semantic memory. In this case, when children engaged in the disambiguation inference, they might not have had access to the heuristic that requires remembering a labeling event (i.e., these objects have been labeled; therefore, the unfamiliar label must refer to the unfamiliar object). If so, in making disambiguation inferences, children must rely on the process of elimination that is based on comparing the novel label with the labels of the previously labeled objects. Such a process may have required substantially more working memory resources and, thus, resulted in a higher proportion of errors.

Overall, the results of Experiments 1 and 2 establish that by 3.5–6.5 years of age, children are not just using speakers' history of accuracy to make predictions about their future information. They are flexibly using accuracy information to reevaluate their beliefs about the past information provided by those speakers.

The lack of difference between the accurate and inaccurate speaker conditions in Experiment 2 suggests that the effect of condition we observed in Experiment 1 was not explained by the increased memory demands associated with processing incorrect labels in the inaccurate speaker condition. Rather, we suggest that children made retrospective inferences about the novel information provided by those speakers after receiving subsequent information about the speakers' reliability.

The pattern of errors in Experiment 1 also supports this interpretation. If incorrect responses were related to increased memory demands for processing incorrect information, then children should have been equally likely to confuse object–label pairings and to choose the distractor object during test. Children, however, were significantly more likely to choose the distractor. It is likely that their choice of object was informed by retrospective inferences (the speaker is unreliable and must have provided wrong information; thus, a completely unfamiliar object is more likely to be the correct referent of the label than either of the previously labeled objects).

One concern with this conclusion is that children might not have made retrospective inferences about the quality of the novel labels provided by the speaker but instead formed domain-general retrospective distrust in an informant's knowledge. That is, instead of making an inference about the speaker's reliability in the verbal domain based on her accuracy or inaccuracy in labeling, children may have made a broader valence inference about the speaker's competence and concluded that she could not be relied on in more general terms. Although our experimental design does not distinguish between these possibilities, there is limited evidence that children do make sweeping inferences about informants' competence (e.g., "halo" and "pitchfork" effects). But in the cases where children have demonstrated halo effects (preferring stronger informants to weaker ones, kinder informants to meaner ones, etc.), the informants were puppets rather than adult humans (e.g., Brosseau-Liard & Birch, 2010; Fusaro, Corriveau, & Harris, 2011). When human adults acted as informants, preschoolers did not demonstrate halo effects (e.g., Koenig & Jaswal, 2011; Sobel & Corriveau, 2010). Given that in our experiments children observed adult human actors and, most likely, had previously formed an association between adults and competence, it is unlikely that by observing three instances of mislabeling they inferred domain-general incompetence or unreliability of the speakers. Future research should investigate this theoretical possibility.

Our findings from Experiment 1 are broadly consistent with the idea that selective word learning during the preschool years might be driven by semantic blocking (Mangardich & Sabbagh, 2018). This process describes the suppression of establishing semantic mapping between a label and its referent, if the label was provided by an unreliable informant, while retaining episodic memory of the labeling event. Children's depressed test performance in the inaccurate speaker condition suggests that despite equivalent conditions of the initial encoding of novel labels, subsequent inaccuracy in labeling interfered with semantic consolidation. Chance-level performance on the disambiguation trials, however, suggests that semantic blocking alone cannot fully explain our findings. If children's retrospective learning was driven by a process through which semantic consolidation, but not episodic memory, was interfered with by the exposure to inaccurate labeling, then we would expect to observe above-chance performance on disambiguation trials in the inaccurate speaker condition. Despite failed semantic consolidation, children should still remember that two of three objects have already been labeled and, therefore, would have selected the object corresponding to the unfamiliar label. The observed chance-level performance on these trials suggests that children did not simply block recently learned semantic mappings after observing their source mislabel familiar objects but instead allowed for the possibility that the previously labeled objects may have different names.

The current results are also consistent with children's general capacity to retrospectively revise their beliefs. When presented with initially ambiguous data, preschoolers can integrate new data to come to appropriate causal conclusions (e.g., Bonawitz, Fischer, & Schulz, 2008; Griffiths, Sobel, Tenenbaum, & Gopnik, 2011; Sobel, Tenenbaum, & Gopnik, 2004). Moreover, the current results are consistent with the "wait and see" strategy for mutual exclusivity (e.g., Savage & Au, 1996). Children potentially keep in mind possible labels for an object for a period of time until the information that allows them to revise their hypotheses becomes available. As Scofield and Behrend (2008) correctly pointed out, children often make these inferences from speakers with whom they have no experience (e.g., Markson & Bloom, 1997), and such a strategy is consistent with previous findings suggesting that children are influenced by pragmatic cues generated by the speakers (Diesendruck et al., 2010; Sobel et al., 2012).

To conclude, our results highlight the role of children's active involvement in their selective learning and add to the growing body of literature on the development of its underlying mechanisms. Earlier work shows that children as young as 18 months already incorporate social cues into their selective learning (e.g., Luchkina, Sobel et al., 2018). By 3 or 4 years of age, children's selective learning is predominantly driven by inferences about speakers' epistemic competence, but there is still a residual effect of associative generalizations (Hermes, Behne, & Rakoczy, 2018; Luchkina, Morgan, & Sobel, 2018). The current findings suggest that by the time children are 3.5–6.5 years old, they make inferences about epistemic competence both prospectively and retrospectively.

Acknowledgments

This research was funded by National Science Foundation (NSF) Grant 1652224 to K.H.C. and NSF Grant 1661068 to D.M.S. We thank Jenna Alton, Kelly Cui, Emily Palmer, and Amita Sekhon of Boston University and Zoe Finiasz and Jayd Heilbrun of Brown University for assistance with participant recruitment, data collection, and coding.

References

Barry, R. A. (2016). Selective word learning in toddlers: An eye tracking investigation of the mechanism. Doctoral dissertation, University of California, Davis. Retrieved from ProQuest Dissertations (accession no. 10165848).

Bernard, S., Proust, J., & Clément, F. (2014). The medium helps the message: Early sensitivity to auditory fluency in children's endorsement of statements. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01412.

Bonawitz, E. B., Fischer, A., & Schulz, L. (2008, January). Training a Bayesian: Three-and-a-half-year-olds' reasoning about ambiguous evidence. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 837–842). Austin, TX: Cognitive Science Society. <https://escholarship.org/uc/item/5mx3g61x>.

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy, 18*(Suppl. 1), E46–E68.

Brosseau-Liard, P. E., & Birch, S. A. (2010). "I bet you know more and are nicer too!": What children infer from others' accuracy. *Developmental Science, 13*, 772–778.

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language, 19*, 360–379.

Diesendruck, G., Carmel, N., & Markson, L. (2010). Children's sensitivity to the conventionality of sources. *Child Development, 81*, 652–668.

Doebel, S., Rowell, S. F., & Koenig, M. A. (2016). Young children detect and avoid logically inconsistent sources: The importance of communicative context and executive function. *Child Development, 87*, 1956–1970.

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology, 110*, 561–574.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35*, 1407–1455.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.

Henderson, A. M., Graham, S. A., & Schell, V. (2015). 24-Month-olds' selective learning is not an all-or-none phenomenon. *PLoS One, 10*(6) e131215.

Hermes, J., Behne, T., & Rakoczy, H. (2018). The development of selective trust: Prospects for a dual-process account. *Child Development Perspectives, 12*, 134–138.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Koenig, M. A., & Doebel, S. (2013). Children's understanding of unreliability. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 235–240). New York: Oxford University Press.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development, 82*, 1634–1647.

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology, 46*, 815–826.

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2008). The role of preschoolers' social understanding in evaluating the informativeness of causal interventions. *Cognition, 107*, 1084–1092.

Luchkina, E., Morgan, J. L., & Sobel, D. M. (2018). Questions answer questions about the mechanism of selective word learning. Manuscript submitted for publication. doi: 10.31219/osf.io/k7qsj.

Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science, 21* e12663.

Mangardich, H., & Sabbagh, M. A. (2018). Mechanisms of selective word learning: Evidence and implications. In M. M. Saylor & P. A. Ganea (Eds.), *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms* (pp. 147–166). Cham, Switzerland: Springer.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*, 813–815.

Nurmsoo, E., & Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused. *Child Development, 80*, 23–27.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology, 126*, 182–192.

Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers. *Cognition, 112*, 415–422.

Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers?. *Journal of Child Language, 30*, 905–924.

Savage, S. L., & Au, T. K. F. (1996). What word learners do when input contradicts the mutual exclusivity assumption. *Child Development, 67*, 3120–3134.

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*, 278–290.

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*, 669–679.

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120*, 779–797.
Sobel, D. M., Sedivy, J., Buchanan, D. W., & Hennessy, R. (2012). Speaker reliability in preschoolers' inferences about the meanings of novel words. *Journal of Child Language, 39*, 90–104.
Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*, 303–333.
Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science, 17*, 443–451.