

Sensitivity of 24-Month-Olds to the Prior Inaccuracy of the Source: Possible Mechanisms

Melissa A. Koenig
University of Minnesota, Twin Cities

Amanda L. Woodward
University of Maryland, College Park

Three studies examined 24-month-olds' sensitivity to the prior accuracy of the source of information and the way in which young children modify their word learning from inaccurate sources. In Experiments 1A, 2, and 3, toddlers interacted with an accurate or inaccurate speaker who trained and tested children's comprehension of a new word–object link. In Experiment 1, children performed less systematically in response to an inaccurate than to an accurate source. In Experiments 2 and 3, after toddlers' comprehension of the new word–object links was tested by the original source, a second speaker requested the target objects. In Experiment 2, children responded randomly in response to the second speaker's requests when novel words were previously presented by an inaccurate source. In Experiment 3, toddlers responded randomly in response to both speakers in the inaccurate condition when their memory for words was taxed by a brief delay period. Taken together, these findings suggest that toddlers attend to accuracy information, that they treat inaccuracy as a feature of a particular individual, and that the word–object representations formed as a result may be fragile and short lived. Findings are discussed in terms of possible mechanisms by which children adjust their word learning from problematic speakers.

Keywords: toddlers, selective learning, word learning, selective trust, testimony

Much of what we know we learn through communication with others. We rely on the reports of others for discoveries in science, the geography of the world, the names of objects, the date of our birth, the name of our town, the identity of our parents, and many other domains that play important roles in our lives. Our deep dependence on human communication and other forms of cultural transmission brings with it central questions about the process by which we acquire knowledge from other people. What cognitive safeguards might protect against false or deceptive input? Although much evidence demonstrates that preschool children are sensitive to variation among informants in knowledge and reliability (Birch, Vauthier, & Bloom, 2008; Clément, Koenig, & Harris, 2004; Corriveau, Meints, & Harris, 2009; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007; Sabbagh & Baldwin, 2001), the developmental origins and nature of this selective trust remain unknown.

A common notion is that children are credulous sponges who unreflectively assent to whatever they are told (Coady, 1992; Gilbert, 1991; Reid, 1764/1970). For Reid (1764/1970), credulity

is a “gift of nature” and is “strongest in childhood” (p. 197). In Reid's view, testimony is treated like firsthand experience; we accept what we are told, just as we accept the “testimony of our senses” or the “testimony of our memory.” According to Wittgenstein (1969), “a child learns there are reliable and unreliable informants much later than it learns the facts which are told it” (Sec. 143). It is not difficult to find more contemporary support for this view, especially with reference to language learning. For Fricker (2006), “the initial stages of language acquisition by a child inevitably occur through a process of simple trust in its teachers—parents and other carers” (p. 225). For Williams (2002), “language learning must in the first instance be conducted in circumstances of a pre-reflective openness or, . . . primitive trust” (p. 49). According to Quine and Ullian (1978), “a force for credulity can be found in the very mechanism for language learning. . . . We learn to use expressions in the circumstances in which we find them used by others” (p. 53). In sum, common sense and scholarship seem to converge in holding that very young learners generally accept any and all conveyed information without reflection.

Recent evidence suggests that by 3 years of age, children cannot be aptly characterized as indiscriminately trusting toward other people. For instance, preschoolers do not learn words from a speaker who shows signs of ignorance regarding their correct meanings (Koenig & Harris, 2005; Sabbagh & Baldwin, 2001), and they choose to accept proper nouns only from speakers who personally know the individuals they are naming (Birch & Bloom, 2002). Furthermore, children of this age systematically track differences between speakers in terms of their past reliability and make persistent use of this source information to selectively learn new words as well as new object functions (Birch et al., 2008; Koenig et al., 2004). Preschoolers make use of information about speakers' past reliability to predict the future assertions of such

Melissa A. Koenig, Institute of Child Development, University of Minnesota, Twin Cities; Amanda L. Woodward, Department of Psychology, University of Maryland, College Park.

Support for this research was provided by National Institute of Child Health and Human Development Grant 5R03HD47432-2 to Amanda L. Woodward and Melissa A. Koenig. We thank Kat Frederick and Elizabeth Hallinan for their lab assistance in this research as well as all of the families who participated.

Correspondence concerning this article should be addressed to Melissa A. Koenig, Institute of Child Development, University of Minnesota, Twin Cities, Minneapolis, MN 55455. E-mail: mkoenig@umn.edu

speakers (Koenig & Harris, 2005), to revise word–object links retrospectively (Scofield & Behrend, 2008), and to make subtle discriminations between speakers who are less than consistent in their accuracy or inaccuracy (Pasquini et al., 2007). A speaker's credibility also appears to influence children's causal learning (Kushnir, Wellman, & Gelman, 2008), their imitation of others' behaviors (Rakoczy, Warneken, & Tomasello, 2009), and their categorization decisions (Jaswal, 2004; Jaswal & Malone, 2007). Finally, these source discriminations continue to guide learning after a day (Scofield & Behrend, 2008) or a week has passed (Corriveau & Harris, 2009), suggesting that children may form relatively stable source representations in these contexts. Taken together, these findings suggest that children consider the knowledge states of their informants and use that information to adaptively learn from reliable or knowledgeable sources.

Although the evidence above demonstrates an adaptive avoidance of problematic sources among preschool children, much less is known about very early word learners. Infants recognize and deny false claims by 16 to 18 months of age (Koenig & Echols, 2003; Pea, 1982); however, this evidence does not tell us whether they are capable of questioning the truth of a novel claim or characterizing an individual speaker as problematic. There is much research demonstrating that, by 16 months of age, infants make powerful use of various cues to infer a speaker's intentions and to determine the meaning of adult acts of reference (Baldwin, 1993; Baldwin & Moses, 1996; Bloom, 2000; Tomasello & Barton, 1994; for reviews, see Baldwin & Moses, 2001; Sabbagh & Baldwin, 2005). Indeed, Baldwin and Moses (2001) argued that such cues are likely to be required in order for robust word learning to occur. These considerations raise an important question about a young word learner's ability to resist learning from unreliable speakers, in spite of the many cues such speakers present in favor of a word–object link. That is, if very young children are given reason to doubt the reliability of a speaker, can they discount the otherwise compelling referential cues that typically promote the learning of a novel word? Or does a person's history of inaccuracy have little influence on children's appeal to these cues? A primary aim in the current studies was to examine whether even very young word learners evaluate the prior accuracy of an informant and use that information when interpreting new messages.

If the prior unreliability of a speaker does have an effect on word learning in young children, our next primary question is why? There are at least two important candidates to consider when investigating the form this mechanism might take, both of which might contribute to selective learning. One possibility is that when given clear signs that a speaker is not reliable, children create stable source representations for that individual that guide later selective learning. This possibility finds indirect support from the finding that 3-year-olds, as well as 4-year-olds, continue to display selective trust several days and even 1 week after originally encountering the accurate and inaccurate informants (Corriveau & Harris, 2008). There is evidence, among adults, that source memories for faces associated with a history of cheating are better than memories for faces associated with trustworthy or irrelevant behavior (Buchner, Bell, Mehl, & Musch, 2009). On this account, children's ability to keep track and filter information selectively may be supported and guided by relatively stable representations that mark the unreliable source.

A second possibility is that resistance to learning from inaccurate sources is at least initially mediated by more immediate responses to inaccurate speakers. That is, children may only weakly encode information from inaccurate sources, filter out word–object links that are trained by inaccurate speakers, and only briefly retain information about the source. Support for this possibility comes from recent research by Sabbagh and colleagues (Sabbagh & Shafman, 2009; Sabbagh, Wdowiak, & Ottaway, 2003). For example, Sabbagh and Shafman (2009) recently showed that 4-year-old children, shortly after hearing novel words from ignorant speakers, gave correct responses to questions about the labeling episode (i.e., "Which one did I say is the blicket?") but not to standard comprehension questions (i.e., "Which one of these things is the blicket?"). When the children were tested again after a 4-min delay, their memories for the ignorant speaker's labeling episode ("Which one did I say is the blicket?") also declined rapidly. On this account, children's attenuated response to problematic speakers occurs because children form less robust semantic representations from unreliable speakers, representations that degrade quickly.

In the current paper, we address two primary questions: Do very young word learners avoid learning words from inaccurate speakers, and, if so, what is the basis of this avoidance? In studies with older children, two informants provide conflicting testimony and children are asked to make a comparison between the two speakers (Birch et al., 2008; Jaswal & Neely, 2006; Koenig et al., 2004; Pasquini et al., 2007). This procedure has proven very informative; however, children's preference for the more accurate of two informants leaves open questions of mechanism and how children manage information from a single unreliable source. In Experiment 1A, by comparing the rates at which children correctly selected the target in response to a single accurate versus inaccurate labeler, we investigated whether 24-month-olds can curtail their word learning from a previously inaccurate source. In Experiments 1B and 2, we explored two possible low-level explanations for this effect (in 1B, confusion regarding a bizarre labeling context; in 2, failure to attend to the inaccurate speaker). In Experiment 3, we directly investigated the possibility that children's memories for the inaccurate speaker's labels are fragile and short lived.

In order to reduce the demands of the current task and present a more sensitive assessment to this age-group, we included two types of novel labels: first labels (for novel objects) and second labels (for familiar objects). As discussed above, discounting the information provided by a lone inaccurate speaker requires resisting the force of otherwise reliable pragmatic cues. We suspected that, especially in the case of first labels, this might stretch the resources of young word learners. Second labels for objects, however, present toddlers with familiar objects whose names are already registered in their vocabularies, and this prior knowledge might allow children to resist second labels from an inaccurate source (Liittschwager & Markman, 1994).

A further important issue in this research is the question of individual differences. We explored, as a first step, the potential role played by variation in infants' vocabulary knowledge. In the domain of object naming, recognizing a claim as false depends on knowing the meaning of the words in question and identifying violations of use. Given that one cannot identify a statement as false unless one has the vocabulary to do so, the child's arsenal for

identifying claims as mistaken likely increases with the child's vocabulary knowledge. Willingness to accept novel information from an inaccurate source may be influenced by the experience the child has had applying his or her knowledge when identifying false claims and their contexts. To investigate this possibility, we obtained a measure of infants' productive vocabulary size by asking parents to complete the short-form version of the MacArthur Communicative Development Inventory (Fenson et al., 2000).

Experiment 1A

Our purpose in Experiment 1A was to provide a preliminary examination of 2-year-olds' sensitivity to past inaccuracy in communication. In a live interaction, children were familiarized to a single informant who labeled three familiar objects accurately or inaccurately. The speaker then introduced children to a newly named target object and an unnamed distractor object. Afterward, the speaker asked children to select the target from among these two objects. These three steps—familiarization, novel label training, and test—were repeated in a second block of trials. If 24-month-old children assent to whatever adults tell them, they may not evaluate novel labels differently when arising from accurate versus inaccurate speakers. Perhaps such young children will accept all claims as true, despite prior inaccuracies. On the other hand, if children are sensitive to the possibility of mistakes, attention to prior accuracy may provide even toddlers with a basis for doubt.

Method

Participants. The participants were sixty-four 24-month-old infants (mean age = 24 months, 2 days; range = 22 months, 26 days to 26 months, 0 days) from the Chicago area. Thirty-two infants participated in the accurate condition (mean age = 24 months, 2 days; 18 boys), and 32 infants participated in the inaccurate condition (mean age = 24 months, 2 days, 17 boys). Parents were contacted by advertisements or mailings and were given \$10 to cover their travel expenses. All participants came from primarily English-speaking environments. Seven additional children participated but were excluded from the final sample. One infant did not complete all trials, and three infants did not comprehend words for one or more of the experimental items, based on parental report. Other reasons for exclusion were experimenter error (2) and side bias (1). The sample of infants was 53% Caucasian, 22% African American, 15% Hispanic, and 9% Asian.

When parents and their infants arrived, an interviewer administered the short-form version of the MacArthur Vocabulary Checklist, Level II (Fenson et al., 2000), a measure for assessing the productive language of children 16–30 months of age. The vocabulary levels of infants in the accurate condition ($M = 48.3$ words, $SD = 2.6$, range = 15–96) and inaccurate condition ($M = 50.6$ words, $SD = 23.5$, range = 6–92) did not differ, $t(62) = 0.406$, *ns*. Participants were divided into two vocabulary groups on the basis of median split. For simplicity, we refer to these groups as “high vocabulary” ($M = 68.8$ words, $SD = 13.4$, range = 50–96) and “low vocabulary” ($M = 30.1$ words, $SD = 10.7$, range = 6–46).

Materials.

Familiarization and test stimuli. The familiarization stimuli included a total of six familiar objects (3 in each block) labeled

accurately or inaccurately depending on condition. After labeling three familiar objects, the speaker went on to provide novel label training for a test item: In one block, one of the test items was a familiar object (e.g., “mido” for shoe; dubbed “second label” because children already know *shoe*) and, in the other block, one of the test items was a novel object (e.g., “blicket” for a black rubber object; dubbed “first label” because the object is novel and does not carry another label). Paired with the target was a distractor object that was given equal verbal attention (i.e., “That’s a nice one”). On first label trials, children were presented with the black rubber object (target in this example) and a colorful woven distractor object and asked, “Where’s the blicket?” On second label trials, children were presented with the shoe (target) and the cup as distractor and asked, “Where’s the mido?” Each of these comprehension tests was presented twice per test pair, for a total of four test trials per participant. This aspect of the procedure (two types of novel label tests) held for Experiments 2 and 3.

Experimental props. Children were tested at a small table in a quiet room on a university campus. On the table was a white oval tray used in presenting the objects. There was also a cardboard box, which facilitated the exchange of objects. It contained a tunnel (the “chute”) and was wrapped in colorful paper. A square aperture on the child’s side of the box allowed children to drop toys inside. The toys fell through the tunnel and came out of the opposite side of the box to the experimenter.

Design and procedure. Parents were given a concrete description of the procedure before the task, were told that children give a range of reactions to false labeling, and were given the option of wearing a sleep mask during the procedure. If they declined, they were asked to refrain from talking or commenting on the items during the experiment. The experimenter began by greeting the child and introducing the task (“Hi! I’d like to show you some things, okay?”). The accuracy of the speaker’s labels was not mentioned during the experiment.

Familiarization. For half of the participants, the experimenter was consistently accurate in her labeling of familiar objects. For example, after placing an object (e.g., a car) on the tray, she pointed to it, touched it, and alternated her gaze between the object and infant, saying, “See this? It’s a car. Look! A car. That’s a car. Can you put the car down the chute for me?” For the other half of participants, the experimenter was consistently inaccurate. For example, after placing the car on the tray, she said, “See this? It’s a duck. Look! A duck! That’s a duck. Can you put the duck down the chute for me?”

Novel label training. After three familiarization trials, children were presented with a pair of objects. One object was labeled: “Look! That’s a blicket. See? A blicket! This is a blicket. Can you put the blicket down the chute for me?” The experimenter gave an equal amount of nonlabeling attention to the distractor object: “Look! See this one? Wow! It’s a nice one. I like this one! Can you put it down the chute for me?”

Familiar and novel label tests. Following the novel label training, the experimenter administered two novel label tests interspersed among three familiar label trials. On novel label test trials, the experimenter placed the two test objects on either side of the tray, looked at the infant, and asked for the previously labeled target object (e.g., “Where’s the blicket? Can you give me the blicket? Can you put the blicket down the chute for me?”) and pushed the tray toward the infant. After the infant made a selec-

tion, the experimenter said “thank you” and proceeded to the next (familiar label) trial. The experimenter administered a second novel test trial, placed the same test objects in a switched position, and requested the target object again (e.g., “Where’s the blicket? Can you show me the blicket? Can you put the blicket down the chute for me?”).

On familiar label trials, which were included to monitor toddlers’ attention in both conditions, the experimenter presented the three prior familiar objects (i.e., car, dog, and horse) in a series of three randomized pairs. Objects were placed on either side of an oval tray, out of infants’ reach, and terms were used accurately. For example, the experimenter placed the dog and horse on either side of the tray and requested one of the objects (e.g., “Where’s the dog? Can you put the dog down the chute for me?”). Inaccurate terms were not used here. If a car was called “duck” during familiarization, car and duck were not paired during testing.

After completing these five test trials (3 familiar comprehension trials and 2 novel label tests), children went on to complete the same steps in a second block of trials. Block B included familiarization (3 more trials of accurate or inaccurate labeling), novel label training, and familiar and novel label test (5 test trials). Thus, at the end of the experiment, each child received a total of four novel label tests and six familiar comprehension trials. How trial blocks A and B were ordered and whether the first set of test objects was novel or familiar was counterbalanced across participants. The identity of the target object, the novel label it was given (“blicket” or “mido”), and the order in which the target versus the distractor were presented were counterbalanced across participants in each condition.

Coding and reliability. Sessions were videotaped and coded offline. Infants’ responses were coded in terms of which of the two test objects was selected first in response to the speaker’s request. Selection was credited to whichever object was indicated first by the child, whether by touching, grabbing, or pointing. If the child reached for both objects, credit was given to the object that he or she put down the chute. If a child hesitated to make a selection, he or she was invited to “show” it (e.g., “Show me/Mommy/Daddy the blicket”). The test phases of 16 randomly selected infants were coded independently by two coders blind to condition. Their choices demonstrated 98.5% agreement.

Results

Preliminary analyses revealed no effects or interactions involving age, gender, block order (Block A vs. Block B first), or trial order (target presented before vs. after the distractor); subsequent analyses collapsed across these factors. Proportion of target choices on the familiar label comprehension trials (e.g., “Where’s the car?”) served as a check of whether infants remained attentive throughout the procedure in both conditions. When asked for familiar items, infants selected the appropriate target 78% of the time in the accurate condition and 76% of the time in the inaccurate condition, $F_s(1, 63) \leq 1$. Thus, when responding to requests for familiar items, infants responded systematically and similarly in the accurate and inaccurate conditions.

The focal analyses tested whether the speaker’s history of accuracy or inaccuracy influenced children’s subsequent propensity to accept a new word from the speaker. To compare the test performance of infants in the accurate versus inaccurate conditions

and to examine the effects of vocabulary knowledge, we calculated a three-way analysis of variance (ANOVA) with condition (accurate vs. inaccurate) and vocabulary knowledge (high vs. low) as the between-subjects variables and novel label type (first vs. second labels) as the within-subjects variable. This ANOVA revealed main effects of condition, $F(1, 60) = 4.17, p = .046, d = 0.481$, vocabulary knowledge, $F(1, 60) = 5.13, p < .027, d = 0.52$, and a trend for a three-way interaction of Label Type \times Condition \times Vocabulary Knowledge, $F(1, 60) = 3.70, p = .059$. The effect of condition results from greater accuracy in response to a previously accurate labeler ($M = 82.1, SD = 26.2$) than to an inaccurate labeler ($M = 69.0, SD = 28.1$).¹ The effect of vocabulary is due to greater overall performance by children with high vocabularies ($M = 82.2, SD = 27.1$) than by those with low vocabularies ($M = 68.3, SD = 26.1$).

To further examine the trend for a three-way interaction, we conducted two-way ANOVAs (with condition as between-subjects variable and label type as within-subject variable) for each of the two vocabulary groups. Toddlers with high vocabularies learned first labels indiscriminately but were less willing to accept second labels from an inaccurate source than an accurate source, as reflected in a Condition \times Label Type interaction, $F(1, 30) = 4.13, p = .051$. In contrast, analysis of toddlers with low vocabularies revealed no effect of condition, $F(1, 30) = 1.64, ns$, and no Condition \times Label Type interaction, $F(1, 30) = 0.80, ns$. Figure 1 illustrates these findings.

Next we assessed whether children were systematically correct in their choices by comparing the proportion of correct choices to chance ($M = 0.50$). Overall, infants in both the accurate condition ($M = 82.1, SD = 26.2$) and the inaccurate condition ($M = 69.0, SD = 28.1$) responded at above chance rates ($t_s > 3.83, p_s < .01$). In the accurate condition, toddlers with high vocabularies performed at above chance levels for first and second labels ($t_s > 5.50, p_s < .001$), whereas in the inaccurate condition, toddlers’ performance was greater than chance for first labels, $t(16) = 5.61, p < .001$, but not second labels, $t(16) = 1.17, ns$. Toddlers with low vocabularies performed better than chance for first and second labels in the accurate condition ($t_s > 2.38, p_s < .05$) and performed at chance for both label types ($t_s < 1.5, ns$) in the inaccurate condition. In sum, these findings suggest a general effect of prior accuracy or inaccuracy on children’s willingness to learn novel terms. It is possible that high-vocabulary children, by learning first labels but not second labels, are employing a different strategy.

Discussion

We examined whether the accuracy of a speaker’s prior labeling affected infants’ evaluation of new information and found that toddlers, as a group, were less likely to correctly identify the referent of a novel word when the labeling source was previously inaccurate. Furthermore, we found that this sensitivity to inaccuracy may be mediated by two additional

¹ An effect for condition was also found among those toddlers who correctly identified all six familiar targets ($n = 22$), $F(1, 18) = 6.54, p = .02$ (accurate condition, $M = 100.0, SD = 0.0$; inaccurate condition, $M = 81.8, SD = 28.3$).

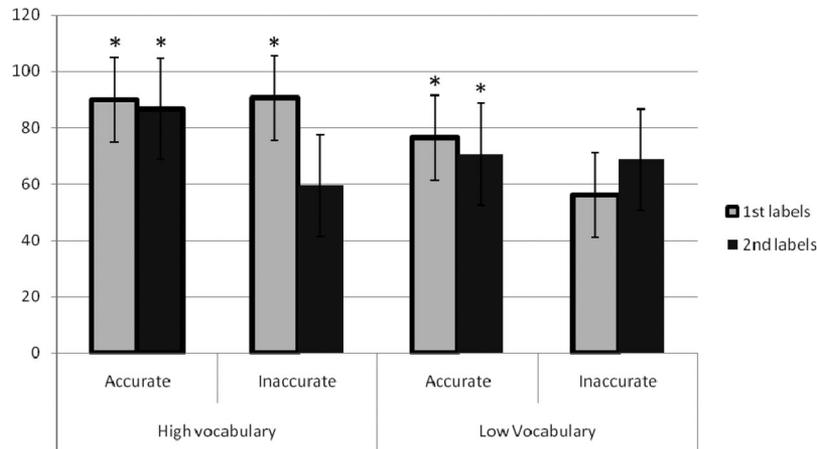


Figure 1. Proportion correct (on first and second label trials) by vocabulary level and condition in Experiment 1A. Error bars indicate standard error. An asterisk indicates values greater than chance ($p < .05$).

factors: variation in vocabulary knowledge and the nature of the information under evaluation (first vs. second label). The vocabulary and label type variables were included in the experiments that follow. We return to discussion of these findings in the General Discussion.

Experiment 1B

In Experiment 1B, our primary aim was to clarify whether the attenuated learning in the inaccurate condition found in Experiment 1A indeed reflected a speaker-specific judgment toward a previously inaccurate source. As discussed earlier, inaccurate speakers present a bizarre mix of cues, and repeated exposure to overt word–object violations may simply cause an undifferentiated confusion in children. We explored this issue in Experiment 1B by having two speakers administer the inaccurate condition: One speaker served as the inaccurate labeler, and a second speaker presented the novel label training and tested children's comprehension (see Table 1). If toddlers' differential responding in Experiment 1A was due to a general confusion caused by false labeling, a context that reproduces that input should result in a similar decrement in performance. On the other hand, if responses reflected speaker-specific judgments

that questioned certain claims made by a particular inaccurate source, children should suspend their doubt and successfully learn second labels from the speaker in Experiment 1B, who had no history of inaccuracy.

Method

Participants. The participants were eighteen 24-month-old infants (mean age = 24 months, 8 days; range = 22 months, 2 days to 25 months, 11 days; 9 female) who were recruited from participant databases in the midwestern United States. None of the children participated in Study 1. Participating families were offered \$10 in appreciation of their time and travel. Children with high and low vocabularies were tested, but given that no effect of condition was found among the low-vocabulary group in Experiment 1A, only the performance of the high-vocabulary toddlers is reported here. Seven additional children participated but were excluded from the final sample because they scored below 50 on the short form of the MacArthur Communicative Development Inventory. The final sample of infants was 69% Caucasian and 31% African American.

Parents completed the short-form version of the MacArthur Vocabulary Checklist, Level II (Fenson et al., 2000). The vocab-

Table 1
Procedural Overview of Experiments 1A, 1B, 2, and 3

Step	1A	1B	2	3
1. Familiarization "That's a dog" ^a	Speaker 1	Speaker 1 ^b	Speaker 1	Speaker 1
2. Novel label training "That's a blicket"	Speaker 1	Speaker 2	Speaker 1	Speaker 1 ^c
3. Test phase Novel label tests: "Where's the blicket?"	Speaker 1	Speaker 2	Speaker 1	Speaker 1
4. Test phase II "Where's the blicket?"			Speaker 2	Speaker 2

Note. Parts 1, 2, and 3 of all experiments were repeated with different stimuli in a second block of trials.
^aFamiliarization utterances were spoken accurately or inaccurately, depending on condition. ^bIn Experiment 1B, Speaker 1 presented inaccurate object labels only. ^cA brief delay period followed the novel object training.

ulary levels of infants in Experiment 1B ($M = 64.6$ words, $SD = 11.1$, range = 50–84) did not differ significantly from those of the high-vocabulary group of Study 1A ($M = 71.1$ words, $SD = 12.6$, range = 50–96), $t(29) = 1.52$, *ns*.

Stimuli. Materials and stimuli were identical to those used in Experiment 1A.

Design and procedure. The procedure was identical to the inaccurate condition of Experiment 1 with one exception: Infants were consecutively presented with two speakers, both female, instead of a single speaker who presented the entire procedure. This procedure included (a) a familiarization phase in which Speaker 1 presented infants with inaccurate labeling of three familiar objects, (b) a novel label training phase in which Speaker 2 introduced a newly named target object and an unnamed distractor object, and (c) a test phase in which Speaker 2 asked infants comprehension questions regarding both the familiar and the novel labels. For clarification, see Table 1. The procedure described below was repeated in a second block of trials.

Familiarization. The first speaker came into the room with a greeting (“Hi! I want to show you some things!”) and used the familiar words (“frog,” “spoon,” “duck”) to incorrectly label the familiar objects (car, dog, horse). After three familiarization trials, she left the room.

Novel label training. A second speaker then entered and said, “Hi! I want to show you some more things!” Speaker 2 proceeded to introduce the two novel test objects. As in Experiment 1A, the target object was labeled four times (e.g., “That’s a blicket!”) and the distractor object was given an equal amount of nonlabeling attention (e.g., “That’s a nice one!”).

Familiar and novel label tests. Following the novel label training, Speaker 2 administered three familiar label comprehension trials along with two novel label test trials, as described in Experiment 1A. Following the final test trial, Speaker 2 left the room. Speaker 1 reentered the room, and the above procedure was repeated for Block 2. As no effects of block order were found in Experiment 1A, block order was held fixed (i.e., Block A first, Block B second).

Results

Familiar label comprehension trials. Consistent with the results of Study 1, infants systematically chose the appropriate object on familiar label trials. When asked for the car, horse, dog, frog, spoon, and duck, they selected the correct object at a level significantly higher than would be predicted by chance ($M = 82.6$, $SD = 22.4$), $t(15) = 5.78$, $p < .001$. No differences were found when comprehension performance was compared to high-vocabulary infants in the false condition of Study 1 ($M = 78.0$, $SD = 21.1$), $t(29) = -0.43$, *ns*.

Novel label test trials. To examine whether children learned both first and second labels, we conducted a one-way ANOVA with first versus second labels as the within-subjects variable and found no effect of label type, $F(1, 17) = .416$, *ns*. Children selected the target at above chance levels for both first labels, $t(17) = 5.25$, $p < .001$, and second labels, $t(17) = 4.95$, $p < .001$. As shown in Figure 2, planned comparisons with Experiment 1A revealed that children with high vocabularies indicated the referents of second labels with greater accuracy in response to the second speaker of Experiment 1B ($M = 86.6$, $SD = 29.9$) than to the inaccurate labeler of Experiment 1A ($M = 61.8$, $SD = 41.3$), $F(1, 34) = 4.10$, $p = .051$, whereas they performed equally well on first label trials in Experiment 1B ($M = 81.8$, $SD = 25.2$) and the inaccurate condition of Experiment 1A ($M = 88.2$, $SD = 28.3$), $F(1, 34) = .73$, *ns*. Furthermore, there was no difference in performance between Experiment 1B and the accurate condition of Experiment 1A ($M = 82.1$, $SD = 26.2$), $F(1, 31) = .61$, *ns*. Children of Experiment 1B (12/18) were more likely to exceed chance performance (4/4 correct) on their own than were children in the inaccurate condition of Study 1 (10/32), in spite of the presence of inaccurate labeling in both contexts, $\chi^2(1, N = 52) = 4.17$, $p = .041$.

Discussion

Experiment 1A showed that children learned less well from an inaccurate than an accurate source. Experiment 1B clarified that

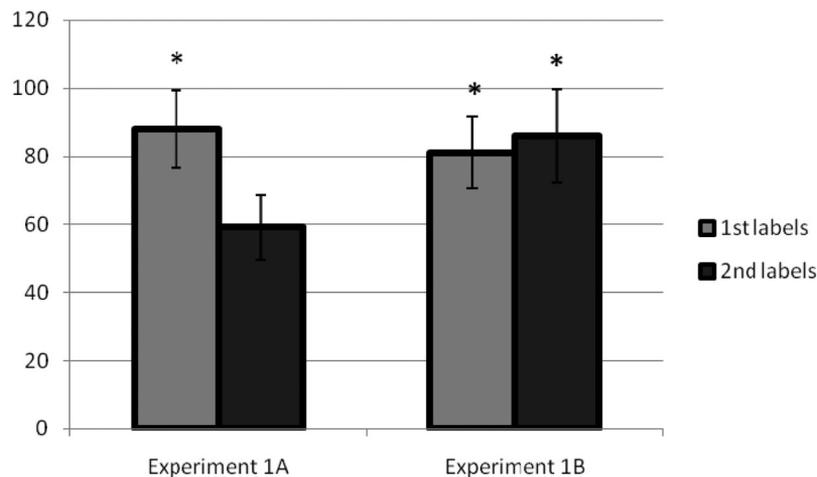


Figure 2. Proportion correct on first and second label trials among high-vocabulary toddlers in the inaccurate condition of Experiment 1A and Experiment 1B. Error bars indicate standard error. An asterisk indicates values greater than chance ($p < .05$).

this result was not likely due to general confusion caused by false labeling but to children's sensitivity to the labeling history of a particular source of information. In spite of the presence of false rather than true labels in Experiment 1B, children responded by selecting targets at a level similar to that found with an accurate labeler. These results speak against the possibility that confusion caused by false labeling can explain the response pattern found in Experiment 1A. Rather, the responses of 2-year-olds toward the inaccurate source likely stem from their attention to the identity of the inaccurate source and their unique treatment of novel information from her. In other words, 2-year-olds treat prior inaccuracy appropriately (i.e., as a feature of a particular individual).

In Experiment 1A, why did toddlers learn first labels equally well from accurate and inaccurate labelers? Although children's learning was, in general, less successful in the inaccurate than the accurate condition, children with high vocabularies responded systematically to the inaccurate source on first label trials. Before possible explanations are discussed, it is important to consider whether the past inaccuracy of the speaker, as presented in Experiment 1A, was a sufficiently strong cue. Although the speaker was unambiguously incorrect in her false labeling of familiar objects, there was a part of the testing procedure that may have rendered the competence of the inaccurate source ambiguous. Recall that during the test phase children were presented with familiar label comprehension checks, which were included to monitor attention. On these trials, the speaker presented two familiar objects (e.g., spoon and frog) and asked for one of them ("Where's the frog?"). Inaccurate terms were not used in these requests. Thus, although the inaccurate speaker was inept when labeling, she was quite competent when requesting objects. As a result, these sensible requests may have degraded the speaker's profile of inaccuracy and increased the cognitive effort required to resolve the speaker's communicative intentions. Experiment 2 was designed, in part, to eliminate such ambiguity.

Experiment 2

One purpose in Experiment 2 was to present toddlers with unmistakably inaccurate sources in an otherwise unchanged procedure. To do this, we simply removed the familiar comprehension test trials described above. In Experiment 2, toddlers experienced familiarization with an accurate or inaccurate source, novel label training, and novel label test in two blocks of trials. We again divided children into high- and low-vocabulary groups to further investigate the role of vocabulary knowledge.

Our second purpose in Experiment 2 was to better understand the cognitive mechanisms that might guide learning from inaccurate sources. The fact that some children responded systematically on first label trials in the inaccurate condition of Experiment 1A suggests that children may indeed have had some trace of a word-object link that was trained by an inaccurate source. In Experiment 2, we asked whether that trace would be sufficiently robust to guide their behavior when they were tested both by the inaccurate source and by a second speaker, whose reliability was not in question. Toddlers were first trained and tested on novel word-object mappings by an inaccurate or accurate source. Then their comprehension was tested a second time by a speaker who was not present at the time of naming. If children indicate the target when tested by the original inaccurate source, this would

suggest that children do encode information from problematic sources and are not diverting their attention away from them. However, if they do not indicate the target in response to a new speaker who follows the inaccurate source, it may be due to the cautious extensions they give to terms presented by inaccurate sources or to the weak semantic memories that support such terms (Sabbagh & Shafman, 2009).

Method

Participants. The participants were forty 24-month-old infants (mean age = 24 months, 3 days; range = 22 months, 10 days to 25 months, 15 days; 22 female) who were recruited from participant databases in the midwestern United States. The only exclusion criterion was that all participants be exposed to English as their dominant language. Those with more than 20% exposure to a language other than English were excluded. Participating families were offered \$10 in appreciation of their time and travel. Three additional children participated but were excluded from the final sample because they did not complete all trials, and two additional were excluded because, based on parental report, they did not comprehend words for one or more of the experimental items. The sample of infants was 61% Caucasian, 30% African American, and 9% Asian American.

The vocabulary levels of infants in the accurate condition ($M = 53.1$, $SD = 22.2$, range = 28–93) and the inaccurate condition ($M = 56.6$, $SD = 13.5$, range = 18–100) did not differ, $t(38) = 0.593$, *ns*. Participants were again divided into two vocabulary groups based on median split: the high-vocabulary group ($M = 68.5$, $SD = 11.8$, range = 56–100) and the low-vocabulary group ($M = 39.8$, $SD = 10.5$, range = 18–54).

Stimuli. Materials and stimuli were identical to those used in Experiments 1A and 1B.

Design and procedure. The procedure was identical to the format of Experiment 1A except for two changes: familiar comprehension trials were removed from the test phase, and a second speaker retested children on their comprehension of the novel word-object pairings at the end of the procedure. Thus, this procedure included (a) a familiarization phase in which infants were presented with three instances of accurate or inaccurate labeling, (b) a novel label training phase in which the speaker introduced a new word-object pairing along with an unnamed distractor object, and (c) a novel label test phase in which the speaker tested children's comprehension of the word-object links. This three-step procedure was repeated in a second block of trials. Then, (d) the first speaker left the room and a second speaker entered and administered four comprehension tests of the novel labels trained by the first speaker.

Familiarization. Familiarization was identical to that of Experiment 1A. Half of the participants interacted with a consistently accurate labeler, and half interacted with an inaccurate labeler.

Novel label training. Novel label training was identical to that of Experiment 1A.

Novel label tests. Following the novel label training, the experimenter immediately administered two novel label comprehension test trials, identical in presentation to those of Experiment 1A. The only difference was that familiar label comprehensions were not included here. The experimenter placed two test objects (the target and distractor) on either side of the tray and asked for the

previously labeled target object (e.g., “Where’s the blicket?”). After the infant made a selection, the experimenter placed the same test objects in a switched position and requested the target object again (e.g., “Can you show me the blicket?”). After completing these two test trials, children went on to complete the same three-step procedure in a second block of trials. Upon completion of the second block and a total of four comprehension test trials, the first experimenter left the room.

Second speaker tests. A second speaker promptly entered and greeted the child (“You’re doing a great job! I’d like to ask you a few more questions.”). The second speaker then placed the test objects (the target and distractor) on either side of the tray, asked for the previously labeled target object (e.g., “Where’s the blicket? Can you give me the blicket? Can you put the blicket down the chute for me?”), and pushed the tray forward. After the child made a selection, the second speaker switched the positions of the objects and requested the target object again (e.g., “Where’s the blicket? Can you show me the blicket? Can you put the blicket down the chute for me?”). This sequence was repeated for the second test pair (e.g., “Where’s the mido?”). The second speaker administered a total of four comprehension test trials. The general order of presentation was preserved across the two speakers: If novel test objects were presented first by the first speaker, the second speaker presented these first as well.

Results

Preliminary analyses. Preliminary analyses showed that there were no main effects of sex, nor did sex interact with any of the independent variables of interest. We found no effects of block order (Block A vs. Block B first) or trial order (target labeled first vs. second); subsequent analyses collapsed across these factors.

Comprehension tests. To compare children’s performance on comprehension tests in response to the first and second speakers in the accurate and inaccurate conditions and to examine further the effects of vocabulary knowledge, we calculated a $2 \times 2 \times 2 \times 2$ ANOVA with condition (Accurate vs. Inaccurate) and vocabulary knowledge (High vs. Low) as between-subjects variables and speaker (First vs. Second Speaker) and label type (First vs. Second Label) as within-subjects variables. This ANOVA revealed a sig-

nificant main effect of speaker, a main effect of vocabulary knowledge, and a significant Speaker \times Condition interaction. The main effect of speaker, $F(1, 36) = 12.50, p < .001, d = 0.47$, resulted from children’s greater overall accuracy in response to the first speaker ($M = 76.8, SD = 22.2$) relative to the second speaker ($M = 66.2, SD = 23.0$). The main effect of vocabulary knowledge, $F(1, 36) = 8.37, p = .014, d = 0.76$, resulted from higher overall performance among the high-vocabulary group ($M = 78.6, SD = 18.1$) relative to the low-vocabulary group ($M = 63.8, SD = 20.3$). Here, in contrast to Experiment 1A, we found that neither label type, $F(1, 36) = 1.73, ns$, nor vocabulary knowledge, $F(1, 36) = .99, ns$, interacted with condition.

Most important, a Speaker \times Condition interaction was found, $F(1, 36) = 5.35, p = .027$. As shown in Figure 3, although no effect of condition was observed in response to the first speaker, $F(1, 39) = 0.28, ns$, there was a significant effect of condition in response to the second speaker, $F(1, 39) = 6.60, p = .013, d = 0.81$. In the inaccurate condition, there was a significant difference in children’s performance in response to the first speaker ($M = 75.0, SD = 22.9$) as compared to the second speaker ($M = 57.5, SD = 21.6$), $t(19) = 4.76, p < .001$, whereas in the accurate condition children responded similarly to the first speaker ($M = 78.7, SD = 21.8$) and the second speaker ($M = 75.0, SD = 21.5$), $t(19) = 0.825, ns$.

Comparisons to chance. One-sample t tests confirmed that children responded systematically in both conditions to the first speaker ($t_s > 5.20, p_s < .001$). Children responded systematically to the second speaker in the accurate condition, $t(19) = 5.21, p < .001$, but not in the inaccurate condition, $t(19) = 1.55, ns$.

Discussion

As shown in Figure 3, toddlers responded systematically to the first speaker in the accurate and inaccurate conditions of Experiment 2. How do we reconcile this with the findings of Experiment 1A? One immediate possibility is that children were confused by the mixed signals that the inaccurate speaker of Experiment 1A presented. Perhaps the cognitive load associated with processing ambiguous referential cues led some children to use mutual exclusivity on second label trials, a strategy that was unavailable when first labels were

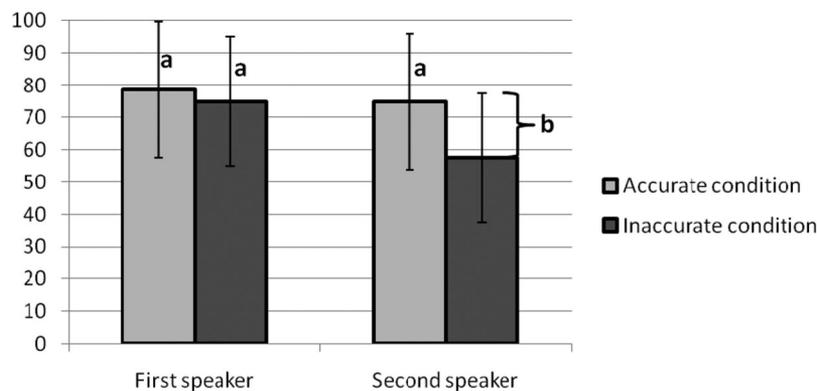


Figure 3. Proportion correct in response to the first and second speaker by condition in Experiment 2. Error bars indicate standard error. a indicates values greater than chance ($p < .05$). b indicates significant condition difference ($p < .05$).

presented (Liittschwager & Markman, 1994). In Study 2, children responded systematically to the inaccurate source when these cues were clarified, which suggests that they could discern the speaker's communicative intent during the novel labeling event in spite of past inaccuracies. Such a result suggests that toddlers are not ignoring the inaccurate source and are not limited by an inability to identify the referential intentions of a previously inaccurate speaker. Furthermore, the distinct response patterns found in response to the first and second speakers are in line with the findings of Experiment 1B in suggesting that toddlers appreciate that inaccuracy is a person-specific feature, which is not extended to others.

The distinct pattern of response across two speakers in the inaccurate condition suggests that, although children attend to the words and cues offered by an inaccurate source, they also note her prior inaccuracies. Perhaps children treated the inaccurate source as presenting idiosyncratic, unconventional word forms that would not likely be shared by another speaker. However, it is also possible that children retained information about the inaccurate speaker's utterances but in a less robust form than that from accurate speakers. Children may have formed very weak representations for forms trained by inaccurate sources, forms that did not survive to support their later interaction with the second speaker. Clarifying the basis for children's attenuated responses to the second speaker was the focus in Experiment 3.

Experiment 3

After they indicated their comprehension to the first speaker, why did children in the inaccurate condition of Experiment 2 not do so in response to the second speaker? We consider two main possibilities here. One possibility is that children encoded the word-object pairings, formed lasting semantic memories for them, but doubted that the second speaker, who showed no signs of unreliability, shared knowledge of the word. Speakers typically use the same words to refer to the same things, and children expect this of them (Clark, 2007; Henderson & Graham, 2005). In Experiment 2, perhaps children limited their assumptions of conventionality or shared knowledge for terms introduced by a marked and inaccurate source. Another possibility is that children encoded the novel word-object pairings sufficiently well to support the initial interaction with the inaccurate speaker but failed to form lasting semantic memories for these terms. In line with what may be children's approach to learning from ignorant speakers (Sabbagh & Shafman, 2009; Sabbagh et al., 2003), children may avoid forming long-term semantic links from speakers who proved inaccurate users of English in the past.

By introducing a 2-min delay between the first speaker's novel label training and comprehension tests, we hoped to clarify the discrepancy in performance between Experiments 1A and Experiment 2 as well as to better understand the kind of representation that toddlers form for new word-object links presented by inaccurate speakers. If children retain lasting word-object links and doubt the conventionality of terms introduced by inaccurate speakers, we expected, Experiment 3 would produce a pattern of results similar to that in Experiment 2: systematic responding to the inaccurate speaker coupled with decreased responding to the second speaker. However, if children form fragile word-object representations from inaccurate sources, the delay between training and test may lead to low responding to the first speaker (and, as a result, to the second speaker).

Method

Participants. The participants were forty 24-month-old infants (mean age = 24 months, 6 days; range = 22 months, 15 days to 25 months, 15 days; 23 female). All participants were exposed to English as their dominant language. Five additional children participated but were excluded from the final sample because they did not complete all trials (3) or failed to know the labels for one or more of the familiarization objects (2). Exact demographic information was not collected, but the database from which they were recruited consists of White middle-class families and includes children of African American, Hispanic, Asian, and Native American parents. Children were tested without regard for racial or ethnic background as long as parents gave consent.

There were comparable numbers of children with high and low vocabularies in each condition. The vocabulary levels of infants in the accurate condition ($M = 50.7$, $SD = 26.3$, range = 12–96) and inaccurate condition ($M = 60.7$, $SD = 23.1$, range = 3–82) did not differ, $t(38) = -1.27$, *ns*. Participants were again divided into two vocabulary groups on the basis of median split: the high-vocabulary group ($M = 76.0$, $SD = 8.5$, range = 57–96) and the low-vocabulary group ($M = 33.3$, $SD = 15.8$, range = 3–55).

Stimuli. Materials and stimuli were identical to those used in Experiments 1A, 1B, and 2.

Design and procedure. The procedure was identical to the format of Experiment 2 with one exception: After the novel label training phase, the child was brought into a neighboring room to play a brief hiding game with a second experimenter. Thus, this procedure included (a) a familiarization phase involving three instances of accurate or inaccurate labeling, (b) a novel label training phase, (c) a 2-min delay period described below, and (d) a test phase in which the child returned and the first speaker tested the child's comprehension of the new word-object links. This four-step procedure was repeated in a second block of trials. Then, (e) the first speaker left the room and a second speaker entered who administered four comprehension tests of the novel labels trained by the first speaker.

Familiarization. Familiarization was identical to that of prior experiments. Half of the participants interacted with a consistently accurate labeler, and half interacted with an inaccurate labeler (E1).

Novel label training. Novel label training was identical to that in Experiments 1A and 2.

Delay period. During a period of approximately two minutes, the child went into an adjacent room and helped a different experimenter hide and find two toys (stuffed frog and bear) in different locations in the room (e.g., behind a chair, in a cabinet).

Novel label tests. Following the hiding game, the child returned to the table and the first speaker (E1) administered two novel label comprehension test trials, identical in presentation to those in Experiment 2. After completing these two test trials, the child went on to complete the same four-step procedure (familiarization, novel label training, delay, test) in a second block of trials. Upon completion of the second block and a total of four comprehension test trials, the first speaker (E1) left the room.

Second speaker tests. A second speaker (E2) promptly entered and administered four comprehension test trials, identical to the second speaker tests of Experiment 2. Again, the general order of presentation was preserved across the two speakers: If the first speaker presented the novel test objects first, the second speaker presented the novel objects first as well.

Results

Preliminary analyses. Because there were no effects of order of presentation, subsequent analyses collapse across block order (Block A vs. Block B first) and trial order (target labeled first vs. second). There were no main effects of sex, nor did sex interact with any of the independent variables of interest.

Comprehension tests. As in Experiment 2, we began by calculating a $2 \times 2 \times 2 \times 2$ ANOVA with condition (Accurate vs. Inaccurate) and vocabulary knowledge (High v. Low) as between-subject variables and speaker (First vs. Second Speaker) and label type (First vs. Second label) as within-subjects variables. The only significant finding from this ANOVA was a main effect of condition, $F(1, 36) = 10.32, p < .001, d = 0.98$ (see Figure 4). Children in the accurate condition ($M = 76.3, SD = 15.1$) performed more accurately overall than did children in the inaccurate condition ($M = 56.3, SD = 24.5$). We found, as in Experiment 2, that neither label type, $F(1, 36) = 0.14, ns$, nor vocabulary knowledge, $F(1, 36) = 2.05, ns$, interacted with condition. Furthermore, in contrast to Experiment 2, no evidence was found for a Speaker \times Condition interaction, $F(1, 36) = 0.16, ns$.

Comparisons to chance. In the accurate condition, one-sample t tests confirmed that children responded systematically to the first speaker ($M = 81.2, SD = 15.9$), $t(19) = 8.75, p < .001$, and to the second speaker ($M = 71.2, SD = 21.8$), $t(19) = 4.34, p < .001$. In contrast, in the inaccurate condition, children responded at chance levels both to the first speaker ($M = 58.7, SD = 23.3$), $t(19) = 1.67, ns$, and the second speaker ($M = 53.7, SD = 32.7$), $t(19) = 0.51, ns$.

Discussion

After the presentation of novel word–object pairings, children of Experiment 3 were engaged in a brief hiding game before their comprehension was tested. As a result, children responded randomly in the inaccurate condition but responded systematically in the accurate condition. In fact, the effect of condition was found to hold both for first speaker tests, $t(38) = 3.22, p = .005$, and for second speaker tests, $t(38) = 1.98, p = .05$. First, these findings help to clarify the results of Experiment 1A and Experiment 2. The

cognitive load of Experiment 1A was likely greater than that of Experiment 2, as manifested in the mixed signals regarding the speaker's reliability and the delay introduced by the familiar comprehension trials. When conditions were simplified and the profile of the inaccurate speaker was clarified in Experiment 2, children encoded both first and second labels from the inaccurate source, and this guided their systematic selections in response to her. Second, the present findings, in concert with those from Experiment 2, suggest that although children do encode information from inaccurate sources (as seen in Experiment 2), the representations formed as a result are weak or temporary and may not survive a short delay (Experiment 3) nor support subsequent interactions (Experiment 2).

General Discussion

Our goals were, first, to investigate whether 24-month-old children modulate their word learning based on a speaker's past reliability and, second, to examine what mechanisms might underlie this adaptive stance. Findings across three experiments converge on the following conclusions: First, 24-month-old children are indeed sensitive to the prior accuracy of the source in a labeling context. This sensitivity was expressed directly, in response to the inaccurate source in Experiments 1A and 3, as well as indirectly, in response to a second speaker in Experiment 2. Second, results from Experiments 1B and 2 converge in demonstrating that children made appropriate, speaker-specific judgments and treated inaccuracy as an idiosyncratic feature of an individual. Third, taken together, the evidence presented here sheds light on the kind of representation that may support children's selective learning. Children's learning was less robust when new terms were trained by the inaccurate speaker than when they were trained by an accurate speaker. When children's memory for the speaker's word–referent link was tested by a second speaker (Experiment 2) or taxed by a brief delay period (Experiment 3), children no longer selected the target object.

Children's differential treatment of accurate versus inaccurate speakers is significant in light of the premium that infants in the second year place on cues to referential intent (Akhtar, Carpenter,

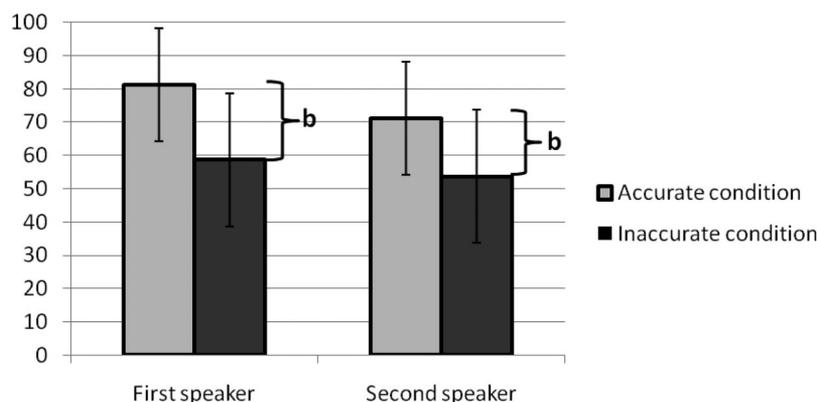


Figure 4. Proportion correct in response to the first and second speaker by condition in Experiment 3. Error bars indicate standard error. a indicates values greater than chance ($p < .05$). b indicates significant condition difference ($p < .05$).

& Tomasello, 1996; Baldwin, 1993; Baldwin et al., 1996; Baldwin & Moses, 2001; Tomasello & Barton, 1994). In the current experiments, infants interacted with a single speaker who routinely presented these powerful cues: She referred to an object with a typical naming phrase while pointing and gazing at the object. When given signs of prior inaccuracy, infants revealed an ability to resist the force of these cues by selecting an object they had never seen labeled. The current results with 2-year-olds further suggest that, even in the absence of a sophisticated theory of mind, children monitor for inaccuracy and fail to learn from an inaccurate source despite the presence of very strong pragmatic cues from that speaker.

By 16 months of age, infants are surprised by and actively refuse claims that conflict with their prior knowledge (Koenig & Echols, 2003; Pea, 1982). Current results suggest that children quickly build on this early sensitivity. By 24 months of age, children not only register the fact that an informant has erred but seem to keep record of that error in a way that affects their subsequent evaluation of new information. Because we did not test children younger than 24 months, we do not yet know when this accuracy tracking begins in the domain of language. However, we do know that extensive evidence of a speaker's inaccuracy is not required. The absence of block effects in any of the studies reported here suggests that toddlers tempered their learning from an inaccurate source after three instances of misnaming. This finding that informants are rapidly appraised for their reliability is consonant with findings with infants (Chow, Poulin-Dubois, & Lewis, 2008), older children (Corriveau & Harris, 2009; Koenig & Harris, 2005), and adults (Ambady & Rosenthal, 1993).

Recall that children selected the target in response to the first speaker of Experiment 2 regardless of her prior accuracy. This result is consistent with the idea that young children, especially at 24 months, may indeed show an immediate trust in what others tell them (Jaswal, Croft, Setia, & Cole, in press), perhaps with aims to share perspective (Tomasello, 2008) or due to a general, unreflective "referential" bias (Csibra, 2003; Csibra & Gergely, 2009). However, the results of this paper further suggest that whether the initial trust leads to robust learning likely depends upon the reliability of the source.

In selective trust research, it is often unclear exactly how children might understand the inaccurate speaker's intentions in these scenarios. After all, an inaccurate speaker presents an odd mix of cues that could simply confuse children: She presents certain referential cues (e.g., conventional English naming phrases, eye gaze, pointing toward the object) and engages in joint attention while, at the same time, she produces false terms. In spite of this bizarre mix, the systematic level of responding shown by toddlers in Experiment 2 suggests that they can interpret the inaccurate speaker's referential cues in a subsequent novel labeling event. In turn, children's current response pattern is not likely due to confusion or uncertainty about how to interpret the naming behavior of an inaccurate source. Instead, it is more likely that the inaccurate speaker is seen as somehow problematic in a way that licenses caution toward her utterances. Given that inaccuracy hinges on the use of unconventional names for objects in these studies, children may be monitoring statements for accuracy per se, evaluating the conventionality of object labels, or combining these approaches.

Experiment 2 provided the first evidence that children, trained on novel terms by an inaccurate source, failed to indicate the targets when subsequently tested by a second speaker. This pattern

stands in contrast to that found in the accurate condition in which children responded systematically when tested by both speakers. First, this pattern, along with that of Sabbagh and Shafman (2009), speaks against the possibility that the circumscribed responses to inaccurate speakers is best explained by children's ignoring or failing to attend to problematic sources. Second, in line with results from Experiment 3, the semantic representations of information from inaccurate sources appear to be very fragile. It may be that anything disruptive can disturb such representations (e.g., the change in speaker of Study 2 and the short delay of Study 3). In Study 3, after a brief distraction, toddlers responded randomly to the inaccurate speaker and failed to retrieve a new word-object link that they had learned only minutes before. Consistent with the "semantic gating" proposal of Sabbagh and Shafman, when inaccurate or ignorant sources attempt to teach children new words, the formation of robust, semantic links is blocked in children as young as 24 months of age.

It is important to note that, while the evidence presented here suggests that children may indeed form weak semantic representations from inaccurate speakers, the nature of the *source* representation remains an important question for future research. For example, Sabbagh and Shafman's (2009) finding that 4-year-olds could not recover their episodic memories after a brief delay raises the possibility that children's memory for the ignorant source may be weak. However, children may not have the support of an explicit understanding when answering questions such as "Which one did I say is the blicket?" but still demonstrate a sustained mistrust of an ignorant source in a selective learning task. Selective trust tasks (Birch et al., 2008; Corriveau & Harris, 2008; Koenig & Harris, 2005), which ask children to choose between a reliable and unreliable informant eight or more times using different information-gathering probes (e.g., "who would you like to ask?" "what do you think it's called?"), are arguably more sensitive than a yes/no method for probing their expectations about source reliability. Taken together, an important possibility raised by these findings is that children's selective learning is based on rapid and enduring appraisals that constitute an example of an implicit bias that children, and perhaps adults, cannot always report on explicitly.

With regard to the issue of individual differences, we examined whether variations in vocabulary knowledge influenced toddlers' ability to trust the more accurate informant. Given the demands of Experiment 1A and the ambiguous status of the inaccurate speaker in that task, perhaps high-vocabulary children were in the best position to deal with that uncertainty by deploying mutual exclusivity and rejecting second labels for objects (Liittschwager & Markman, 1994). Future research should explore this issue, but evidence from Experiments 2 and 3 suggests that the reliability of the speaker affected both high- and low-vocabulary children's subsequent treatment of new words.

These findings raise the possibility that even very young language learners evaluate the quality of verbal information they receive, and this early ability may form the basis for understanding language as a comment on experience. Children who treat information from an utterance as being only provisionally true (or possibly false) can be credited as discriminating between information conveyed in an utterance and the reality it comments on. Recall that on Reid's proposal (1764/1970), young word learners listen to others around them, take in new information uncritically, and update their beliefs, much as they do when they have a new visual experience. Our findings call this

proposal into question at early stages of language development by suggesting that even young children may be capable of questioning the potential truth of a statement.

References

- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635–645.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.
- Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–843.
- Baldwin, D. A., Markman, E., Bill, B., Desjardins, N., Irwin, J., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67*, 3135–3153.
- Baldwin, D. A., & Moses, L. J. (1996). The ontogeny of social information gathering. *Child Development, 67*, 1915–1937.
- Baldwin, D. A., & Moses, L. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309–329.
- Birch, S. A. J., & Bloom, P. (2002). Preschoolers are sensitive to the speaker's knowledge when learning proper names. *Child Development, 73*, 434–444.
- Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018–1034.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behavior, 30*, 212–224.
- Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science, 11*, 761–770.
- Clark, E. V. (2007). Conventionality and contrast in language and language acquisition. *New Directions for Child and Adolescent Development, 115*, 11–23.
- Clément, F., Koenig, M., & Harris, P. L. (2004). The ontogenesis of trust. *Mind & Language, 19*, 360–379.
- Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford, England: Oxford University Press.
- Corriveau, K. H., & Harris, P. L. (2008). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science, 12*, 188–193.
- Corriveau, K. H., Meints, M., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology, 27*, 331–342.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society, London: Biological Sciences, 358*, 447–458.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.
- Fenson, L., Pethick, S., Renda, C., Cox, J., Dale, P., & Reznick, S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*, 95–116.
- Fricker, E. (2006). Testimony and epistemic authority. In J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 225–250). Oxford, England: Oxford University Press.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.
- Henderson, A., & Graham, S. A. (2005). Two-year-olds' appreciation of the shared nature of novel object labels. *Journal of Cognition and Development, 6*, 381–402.
- Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development, 75*, 1871–1885.
- Jaswal, V., Croft, A., Setia, A., & Cole, C. (in press). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*.
- Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*, 263–283.
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17*, 757–758.
- Koenig, M. A., Clément, F., & Harris, P. A. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.
- Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition, 87*, 179–203.
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.
- 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.
- Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology, 30*, 955–968.
- Pasquini, E., Corriveau, K., Koenig, M., & Harris, P. (2007). Children monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216–1226.
- Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two- and three-year olds. *Journal of Child Language, 9*, 597–626.
- Quine, W. V., & Ullian, J. S. (1978). *The web of belief*. New York, NY: McGraw-Hill.
- Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development, 24*, 61–69.
- Reid, T. (1970). *Inquiry into the human mind*. Chicago, IL: Chicago University Press. (Original work published 1764)
- Sabbagh, M., & Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.
- Sabbagh, M. A., & Baldwin, D. A. (2005). Understanding the role of perspective taking in young children's word learning. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds*. Oxford, England: Oxford University Press.
- 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.
- Scotfield, J., & Behrend, D. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*, 278–290.
- Tomasello, M. (2008). *The origins of human communication*. Cambridge, MA: MIT Press.
- Tomasello, M., & Barton, M. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*, 639–650.
- Williams, B. (2002). *Truth and truthfulness*. Princeton, NJ: Princeton University Press.
- Wittgenstein, L. (1969). *On certainty* (D. Paul & G. E. M. Anscombe, Trans.). Oxford, England: Blackwell.

Received August 12, 2008

Revision received December 7, 2009

Accepted December 17, 2009 ■