Research Article

Trust in Testimony

Children's Use of True and False Statements

Melissa A. Koenig, Fabrice Clément, and Paul L. Harris

Harvard University

ABSTRACT—*The extent to which young children monitor and use the truth of assertions to gauge the reliability of subsequent testimony was examined. Three- and 4-year-old children were presented with two informants, an accurate labeler and an inaccurate labeler. They were then invited to learn names for novel objects from these informants. The children correctly monitored and identified the informants on the basis of the truth of their prior labeling. Furthermore, children who explicitly identified the unreliable or reliable informant across two tasks went on to demonstrate selective trust in the novel information provided by the previously reliable informant. Children who did not consistently identify the unreliable or reliable informant proved indiscriminate.*

Any particular piece of testimony might be misleading for a child. Evolutionary biologists have emphasized the deceptive ploys that are likely to emerge in animal signaling systems (Dawkins & Krebs, 1978). In human communication, the dangers are, if anything, even more diverse: The testifier could be making an honest mistake, telling a lie, or not clearly indicating whether the claim is, for example, tentative, factual, or fictional. This observation raises an important empirical question about children's ability to manage such information: Do children have an undifferentiated trust in whatever testimony is offered, or do they display some selectivity in what to believe, whom to trust, and when?

One method children could use to evaluate another person's testimony is to determine whether the claim is consistent with their own past experience. Recent evidence suggests that such caution is present in infancy, at early stages of language acquisition. Koenig and Echols (2003) reported that 16-month-olds directed more attention toward human speakers who falsely labeled objects (e.g., "That's a dog" in reference to a cup) than toward those who truthfully labeled objects. In fact, many infants attempted to correct speakers' false labels through their own pointing and labeling. Similarly, Pea (1982) found that 18-month-olds, with a mean utterance length of 1.0, explicitly rejected false but not true affirmatives (e.g., "That's a ball" in reference to a car) by saying "no." Thus, when human speakers provide information that conflicts with infants' knowledge and experience, infants demonstrate an ability to recognize, correct, and deny assertions that they know to be false.

Address correspondence to Melissa A. Koenig, Harvard Graduate School of Education, 502 Larsen Hall, Appian Way, Cambridge, MA 02138; e-mail: koenigme@gse.harvard.edu.

Such abilities to evaluate questionable testimony extend beyond infancy. Recently, we found that when testimony is pitted against children's first-hand observations, preschoolers generally rely on the latter (Clément, Koenig, & Harris, in press). Three- and 4-year-old children were presented with two puppets, one who named colors accurately and one who named them inaccurately. Subsequently, the children were shown a colored pom-pom that was put inside a box. When the reliable and the unreliable puppet were asked the color of the pom-pom, both stated a color different from what the child had seen. However, when asked the same question, children in both age groups generally ignored the puppets and relied on their prior first-hand observation. Thus, even when presented with testimony from a previously reputable source (i.e., the reliable puppet), children relied on their own observation.

Further evidence that young children reject false propositions comes from studies on deductive reasoning. When children ranging from 2 to 6 years in age are presented with a syllogism such as "All fish live in trees. Tot is a fish. Does Tot live in a tree?" they typically deny that conclusion and justify that claim with reference to their own knowledge: "Because fish live in the water" (Dias & Harris, 1988, 1990; Leevers & Harris, 1999, 2000; Richards & Sanderson, 1999). However, when children are given false premises as part of a story or make-believe context, they are likely to suspend empirical considerations and reason accurately from the supplied premises (e.g., "Yes, Tot lives in a tree"). Thus, children's typical failure to reason from false premises cannot be attributed to any logical inability. Instead, it reflects their unwillingness to treat such premises as a basis for reasoning, an aversion that they inhibit when prompted to regard them as suppositions or make-believe assertions (Harris, 2002).

In sum, when young children have well-established knowledge of a given fact—for example, they know what an object is called, they know what color it is, or they know the properties of the class to which it belongs—they do not accept statements that contradict those known facts. They correct speakers who make false statements and refuse these statements as bases for subsequent reasoning. Although this is an important protective device when children's knowledge and other people's testimony conflict, children often encounter new information that conflicts with nothing simply because they lack information about the topic in question. To gain understanding of, for example, new words, future events, digestion, God, and the afterlife, children presumably depend on information from other people. For an array of

unfamiliar objects and events, the evaluation of the informant becomes essential.

One important marker of an informant's trustworthiness is whether that informant typically spoke the truth in the past. As we have noted, following any given piece of testimony, young children appear to react differently—and appropriately—depending on whether they know an informant's claims to be true or false. Granted that sensitivity, young children might keep track of an informant's reliability and designate particular informants as more or less accurate. The first aim of the present study was to investigate whether 3- and 4-year-old children are able to differentiate between informants in this fashion. To test this issue, we presented children with three familiar objects (a ball, cup, and book) that were each labeled by two different people. One person consistently labeled the objects accurately (e.g., called the book "a book"), and the other person consistently labeled them inaccurately (e.g., called the book "a chair"). Subsequently, the children were asked to identify the person who said something wrong or right. If the children succeeded in making this judgment and correctly identified the inaccurate or accurate labeler, we would have evidence that early sensitivity to the accuracy of content (i.e., the claims themselves) is accompanied by an ability to track the identity of accurate versus inaccurate informants.

A further important question is whether children who can track the accuracy of an informant use that information in judging whether that individual is to be trusted regarding subsequent novel information. More specifically, do children selectively trust a speaker on the basis of the speaker's previous reliability, particularly when they have no prior knowledge of the topic in question? To test this issue, we presented children with a series of three novel objects. The same two speakers who had previously labeled familiar objects accurately or inaccurately provided new labels for the novel objects. For example, the previously inaccurate speaker called a novel object a "mido," and the previously accurate speaker called it a "toma." The children were then asked to say what the object was called. If children treat the accuracy of previous labeling as relevant to a speaker's current reliability, they should be more likely to learn new words from previously accurate than from previously inaccurate informants.

METHOD

Participants

Fifty-three children participated in the study: twenty-six 3-year-olds ($M = 3$ years, 4 months; range = 2 years, 11 months to 3 years, 11 months) and twenty-seven 4-year-olds ($M = 4$ years, 6 months; range = 4 years, 0 months to 5 years, 2 months). Approximately half of the children were female. All were recruited from Harvard University child-care centers in Cambridge, Massachusetts, and were tested there by a single experimenter.

Materials

The visual stimuli were a series of six video clips, corresponding to three familiarization trials and three test trials. Each clip presented the same three actors and a different object. On each video clip, a female actor wearing a red shirt and a female actor wearing a blue shirt sat at a table, and appeared at the left and right sides of the screen. All trials began when the third actor, a male, placed an object on the table. On *familiarization* trials, the objects were familiar: a ball, a cup, and a book. On *test* trials, the objects were unfamiliar: a colorful woven bamboo object; a white rubber object; and a red paper object. The order of trials within the familiarization (ball, cup, book) and test (woven, rubber, paper object) periods was maintained across participants.

Design and Procedure

To introduce the task, the experimenter indicated a picture of the actors and said, "I've got these two friends. See? One has a blue shirt, and one has a red shirt. They're going to show you some things and tell you what they are called. Let's watch." The reliability of the actors' testimony was not mentioned in the introduction.

Familiarization

In each video clip, after placing an object (e.g., a ball) on the table, the third actor asked each of the labeling actors, "Can you tell me what this is called?" The reliable actor consistently responded correctly (e.g., and said, "That's a ball"). The unreliable actor consistently responded incorrectly (e.g., and said, "That's a shoe"). After both of the actors named the object, the children were asked, "Can you tell me what this is called?"

For half of the participants, the red-shirted actor was consistently reliable, and for the other half, the blue-shirted actor was consistently reliable. The seating position (right vs. left side of the screen) of the labelers was counterbalanced within participants.

After familiarization, the children were presented with a total of five test questions of two types: (a) two explicit judgments of an informant's reliability and (b) three inferences regarding the meaning of novel words.

First Explicit Judgment Trial

After the third familiarization trial, the experimenter paused and said, "These people told you about a lot of things." Half of the children were asked to identify the reliable labeler: "Did any of them say something right?" The remaining children were asked to identify the unreliable labeler: "Did any of them say something wrong?" When children said "yes," the experimenter asked, "Who? Point to the person who said something [right/wrong]." When children responded "no," the experimenter corrected them, "Actually, one of them did say something [right/wrong]. Which person said something [right/wrong]?"

Novel Test Trials

Following the first explicit judgment question, participants were presented with three novel test trials, identical in structure to the familiarization trials. The center actor placed a novel object on the table and asked each of the other actors, "Can you tell me what this is called?" In reference to the first novel object, for example, one actor said, "That's a mido." The other actor said, "That's a toma." Children were then asked, "Can you tell me what this is called, a mido or a toma?" See Table 1 for a complete list of names and objects used in both the familiarization and the test trials.

Second Explicit Judgment Trial

After the three novel test trials, the children were asked the same explicit judgment question again, "One of these people kept saying something [wrong/right]. Which one kept saying something [wrong/right]?"

The entire session lasted approximately 5 to 10 min and was typically videotaped.

TABLE 1
Experimental Stimuli

Object	Names
Familiarization	
Ball	Ball, shoe
Cup	Cup, dog
Book	Book, chair
Test	
Colorful woven bamboo object	Toma, mido
White bulbous rubber object	Wug, dax
Red textured paper object	Blicket, dawnoo

Note. One true and one false name were used for each familiarization object. Two nonwords were used as names for each novel test object.

RESULTS

Familiarization Trials

On 97.5% of the familiarization trials, the children provided the appropriate label when asked, "Can you tell me what this is called?" Thus, they were not misled by false labelers in regard to object names that they knew.

Explicit Reliability Judgments

We determined the proportion of explicit judgment questions (out of two) that each child answered correctly. The children performed at above-chance levels on these two questions, with an overall mean of .76 ($SD = .33$), $t(53) = 5.55$, $p < .0001$ (two-tailed). Four-year-olds ($M = .83$, $SD = .28$) performed somewhat better than 3-year-olds ($M = .69$, $SD = .38$); however, the difference between the two age groups was not significant, $t(51) = 1.56$, $p = .126$. Both 3-year-olds, $t(26) = 2.61$, $p < .02$, and 4-year-olds, $t(27) = 5.79$, $p < .0001$, performed at above-chance levels.

Nonparametric tests were used to examine whether performance differed on the first (.79 correct) and second (.72 correct) explicit judgment trials. Binomial tests confirmed that performance was better than would be expected by chance on both the first ($p < .001$) and the second ($p < .005$) question. Moreover, the children were no more likely to err on the second than on the first question. Ten children (four 3-year-olds and six 4-year-olds) responded correctly on the first question and incorrectly on the second. Six children (five 3-year-olds and one 4-year-old) responded incorrectly on the first question and correctly on the second. A McNemar test comparing these two subgroups showed that children were no more likely to err on the second question than the first, $\chi^2(1, N = 16) = 0.56$, n.s.

In summary, children were generally accurate in identifying which of the two informants had said something right (or wrong), both immediately after the familiarization trials and again after the test trials. Occasional errors of identification did occur, especially in the younger age group, but this age difference failed to reach significance. We now consider whether the children used this information in the test trials.

Relation Between Explicit Reliability Judgments and Test-Trial Performance

The children's responses in the novel test trials were examined in relation to their answers to the explicit judgment questions. We compared the test-trial performance of two groups of children: those who responded correctly on both explicit judgment questions and those who responded correctly on only one or neither of these questions.

Each child received a score from 0 to 3, depending on how often he or she correctly indicated the reliable labeler's name for an object. Although responses included references both to words and to speakers, a response was coded as correct only if the child appropriately supplied the correct name or rejected the wrong name. Thus, a response was considered incorrect if the child selected one of the labelers (e.g., "what blue said") or the name given by the unreliable labeler. "Don't know" responses and alternative names for objects were also scored as incorrect.

A total of thirteen 3-year-olds (50%) and nineteen 4-year-olds (70%) responded correctly on both explicit judgment questions. Overall, these 32 children performed at above-chance levels on test trials ($M = .67$, $SD = .29$), $t(32) = 3.22$, $p < .005$. The test-trial performance was significantly above chance within this group for both 3-year-olds ($M = .69$, $SD = .32$), $t(13) = 2.18$, $p < .05$, and 4-year-olds ($M = .65$, $SD = .28$), $t(19) = 2.30$, $p < .05$. A total of 21 children failed to answer both explicit judgment questions correctly. On test trials, these children performed at chance levels ($M = .40$, $SD = .33$), $t(21) = 1.44$, n.s. Three-year-olds who failed to answer both explicit judgments correctly performed at chance levels ($M = .36$, $SD = .32$), $t(13) = 1.59$, n.s., as did 4-year-olds ($M = .46$, $SD = .35$), $t(8) = 0.33$, n.s.

To compare the relative accuracy on test trials of the two subgroups of children (i.e., those who had answered both explicit judgment questions correctly vs. those who had not), we calculated a three-way analysis of variance (ANOVA) with between-subjects factors of explicit judgment accuracy (both correct or 0–1 correct), age (3 or 4 years of age), and wording of the explicit judgment question ("right," "wrong"). Because the test-trial data were proportional, an arcsine transformation was applied to normalize the distribution for analysis. The analysis produced a significant main effect for explicit judgment accuracy, $F(1, 52) = 8.35$, $p < .01$. Children who successfully identified the unreliable or reliable labeler both times performed better on test trials ($M = .67$, $SD = .29$) than those who answered one or more explicit judgment questions incorrectly ($M = .40$, $SD = .33$). There were no other significant main effects or interactions.

Alternative Names and "Don't Know" Responses Excluded

Children's production of alternative names and "don't know" responses may reflect unwillingness to choose between the two alternatives given by the labelers. Hence, it may be inappropriate to classify such responses as "incorrect." Thus, in a separate analysis we examined only those trials (92% of trials) on which children clearly indicated the name provided by either the reliable or the unreliable labeler. This more focused coding strategy resulted in the same pattern of results as that just reported.

Overall Accuracy on Explicit Judgment Questions and Novel Test Trials

Finally, we compared children's accuracy on explicit judgment questions and novel test trials, performing a two-way ANOVA with age (3, 4) as a between-subjects variable and task type (explicit judgment, novel test trial) as a within-subjects variable. Again, an arcsine transformation was applied to normalize the distribution for analysis. There was a significant effect of task type, $F(1, 52) = 14.11$,

$p < .0001$, but no other main effects or interactions. The main effect of task confirms that children of both ages performed better on explicit judgment questions ($M = .76$, $SD = .33$) than on novel test trials ($M = .56$, $SD = .33$). In fact, overall performance on test trials did not differ from chance, $t(53) = 1.31$, n.s.

DISCUSSION

The first aim of the present study was to investigate—via explicit judgment questions—whether children could keep track of the reliability of particular informants over successive occasions. The second aim was to examine—via children's performance on test trials—whether children would use a speaker's prior accuracy to assess whether that speaker should be trusted regarding new information. Finally, we examined whether children would retain information about a speaker's accuracy by asking the explicit judgment question a second time at the end of the experiment.

Given that infants and young children are sensitive to the difference between true and false claims, we anticipated that children would be able to go beyond this sensitivity and link truthfulness (or the lack of it) with the identity of a particular informant. In fact, 3- and 4-year-olds proved very competent at identifying accurate versus inaccurate informants. The finding that preschool-age children evaluate the accuracy of testimony with a critical eye for the identity of the testifier is an important first step in the understanding of children's potential for selective trust in testimony.

Can children who keep track of an informant's prior accuracy use that information to judge whether that individual is to be trusted regarding new information? The answer is yes. A key factor associated with children's trust in the reliable informant was their ability to both identify the unreliable or reliable labeler and retain that person's identity over a short period. In contrast, children who failed to answer both explicit judgment questions correctly demonstrated no selective learning from the reliable labeler. It appears that consistent, perhaps categorical, judgments of a person's prior accuracy enabled children to favor new information from the reliable source.

Such a result is consistent with research demonstrating that children learn words better from knowledgeable than uncertain speakers (Sabbagh & Baldwin, 2001) and is congruent with recent findings by Robinson and Whitcombe (2003), who reported that young children rejected testimony from an informant less informed than they were. Thus, although young children may struggle to identify information sources (Taylor, Esbensen, & Bennett, 1994; Gopnik & Graf, 1988), they are skilled at judging whether information has come from a more or less trustworthy source. Nevertheless, important and interesting questions remain. For example, do children conceive of trustworthiness as a general epistemic trait or one that is tied to a particular type of information? In our procedure, labeling was the only behavior exhibited by the informants. Hence, children may have formed a narrow judgment about the informants' labeling accuracy. Alternatively, they may have formed a more wide-ranging judgment about the informants' accuracy with respect to various types of information: object location, causal explanations, temporal judgments, and so forth.

Recall that the proportion of correct replies was .76 for explicit judgment questions but .56 on test trials. Children may have been able to assess the speakers' accuracy, but using that information to infer which of two novel terms to accept was more difficult. The relatively deflated performance on test trials cannot be explained by memory loss given children's good performance on both the second and the first explicit judgment questions. Indeed, even those children who commented on the speakers, either spontaneously or in response to a postexperimental prompt, performed less accurately on test trials than on the explicit judgment questions. Fifteen children (four 3-year-olds and eleven 4-year-olds) produced at least one such comment on the speakers; examples include, "The one she said [indicating the speaker wearing the red shirt]," "Listen to blue," "The one in the red shirt kept making it right," "The one with the blue shirt—she says correct things . . . not this one [red-shirted actor]—this one says wrong things." Despite their astute observations of the speakers, these 15 children were also less accurate on test trials ($M = .60$, $SD = .36$) than on explicit judgment questions ($M = .90$, $SD = .18$).

Granted that children were able to determine who was the unreliable or reliable informant, why was that knowledge not always used to learn new words? Arguably, children did not take the speakers' previous mistakes seriously. However, this seems unlikely given that the informants' demeanor was serious and children rarely laughed or gave other indications that they found the mistaken behavior funny. Moreover, the terms "right" and "wrong" appeared to be acceptable to the children as terms for describing the informants' behavior, as indicated by the children's own spontaneous comments, in which these terms occurred much more frequently than "silly" and "funny." A more plausible explanation concerns the use of names in our study. We can assume that in line with their intentions, people typically present accurate and conventional descriptions of things to young children (Grice, 1969). Thus, true labeling by adults may be such an entrenched expectation among young children that the false labeler in our study struck our participants as a violation (of linguistic convention) that proved difficult for them to understand and reason upon. If so, we might expect a different pattern of results when children reason about inaccurate and accurate behavior in domains other than adults' language use. We are currently examining this possibility by presenting children with accurate and inaccurate actors in a nonlinguistic task.

The present results also challenge the notion that young children are reluctant to predict consistency in other people's behavior. For example, Kalish (2002) found that evidence of a person sharing with others on one occasion did not lead 5-year-olds to predict future sharing from that person. In contrast, many 3-year-olds in the present study believed that a person who was truthful in the past would be truthful in the future. Different methodologies (e.g., single vs. multiple instances) may be responsible for the difference in results. Future studies asking children to interpret different types of human behavior would contribute toward clarifying such discrepancies.

Finally, we might have expected performance on the second explicit judgment question to be degraded because of memory demands, especially given that the children gained no new information about the speakers' accuracy during test trials. Yet the children's level of accuracy on the second judgment question was similar to their level on the first. If we presented the same informants to children after a longer delay—for example, after 1 to 2 days or a week (Drummey & Newcombe, 2002)—would they recall who was trustworthy and who was not? It is an open question whether an enduring memory for trustworthy individuals would manifest itself in explicit judgments, the uptake of novel information, or both.

Although truth telling among speakers is generally noble and even sensible, unrestricted credulity among listeners is neither noble nor

sensible (Quine & Ullian, 1978; Williams, 2002). The present study shows that listeners as young as 3 years of age can keep their credulity within certain limits. This ability to accept testimony from trustworthy individuals may emerge as part of children's understanding of other people's mental states, or it may surface earlier, for example, at the onset of word learning. Whatever the origins of such trust, these findings show that when young children lack knowledge about a given topic, they have recourse to an important precautionary strategy: attend to the accuracy of what you hear and trust in previously reliable informants.

Acknowledgments—Support for this research was provided by National Institute of Child Health and Human Development Grant F32 HD42860 to M. Koenig. We thank Rita Astuti and Laurence Kaufmann for helpful feedback on earlier presentations of this work.

REFERENCES

Clément, F., Koenig, M.A., & Harris, P.L. (in press). The ontogenesis of trust. *Mind and Language*.

Dawkins, R., & Krebs, J.R. (1978). Animal signals: Information or manipulation? In J.R. Krebs & N.B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). Oxford, England: Blackwell Scientific Publications.

Dias, M., & Harris, P.L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, *6*, 207–221.

Dias, M., & Harris, P.L. (1990). The influence of imagination on reasoning by young children. *British Journal of Developmental Psychology*, *8*, 305–318.

Drummey, A.B., & Newcombe, N.S. (2002). Developmental changes in source memory. *Developmental Science*, *5*(4), 48–62.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.

Grice, H.P. (1969). "Utter's meaning and intentions." *Philosophical Review*, *78*, 147–177.

Harris, P.L. (2002). What do children learn from testimony? In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 316–334). Cambridge, England: Cambridge University Press.

Kalish, C.W. (2002). Children's predictions of consistency in people's actions. *Cognition*, *84*, 237–265.

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.

Leevers, H.J., & Harris, P.L. (1999). Persisting effects of instruction on young children's syllogistic reasoning with incongruent and abstract premises. *Thinking & Reasoning*, *5*, 145–173.

Leevers, H.J., & Harris, P.L. (2000). Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal of Experimental Child Psychology*, *76*, 64–87.

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* (2nd ed.). New York: Random House.

Richards, C.A., & Sanderson, J.A. (1999). The role of imagination in facilitating deductive reasoning in 2-, 3- and 4-year-olds. *Cognition*, *72*, B1–B9.

Robinson, E.J., & Whitcombe, E.C. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74*, 48–62.

Sabbagh, M.A., & Baldwin, D.A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070.

Taylor, M., Esbensen, B.M., & Bennett, R.T. (1994). Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development*, *65*, 1581–1604.

Williams, B. (2002). *Truth and truthfulness*. Princeton, NJ: Princeton University Press.

(Received 7/22/03; Revision accepted 9/8/03)