

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Elizabeth J. Robinson and Erika Nurmsoo

Article Title: When do children learn from unreliable speakers?

Year of publication: 2008

Link to published version: (NYP) Journal website =
<http://www.sciencedirect.com/science/journal/08852014>

Running head: Learning from unreliable speakers

When do children learn from unreliable speakers?

Elizabeth J. Robinson and Erika Nurmsoo

University of Warwick

Author note. The research was supported by a grant from the Economic and Social Research Council, U.K. (RES-000-22-1847).

Correspondence concerning this article should be addressed to E.J. Robinson, Department of Psychology, Warwick University, Coventry, CV4 7AL, U.K. Electronic mail may be sent to e.j.robinson@warwick.ac.uk.

FINAL REVISION Cognitive Development August 11, 2008

Abstract

Children do not necessarily disbelieve a speaker with a history of inaccuracy; they take into account reasons for errors. Three- to 5 year-olds ($N = 97$) aimed to identify a hidden target in collaboration with a puppet. The puppet's history of inaccuracy arose either from false beliefs, or occurred despite his being fully informed. On a subsequent test trial, children's realistic expectation about the target was contradicted by the puppet who was fully informed. Children were more likely to revise their belief in line with the puppet's assertion when his previous errors were due to false beliefs. Children who explained this puppet's prior inaccuracy in terms of false belief were more likely to believe the puppet than those who did not. As children's understanding of the mind advances, they increasingly balance the risk of learning falsehoods from unreliable speakers against that of rejecting truths from speakers who made excusable errors.

When do children learn from unreliable speakers?

Children acquire much of their knowledge about the world from what people tell them. This ability to learn from other people's experiences provides great benefits, but also puts children at risk of learning falsehoods, since even well intentioned informants can be wrong. To manage this risk, children need ways of predicting the likely truth of what they are told. One way children do this is by tracking the speaker's history of accuracy: 4- and 5- year-olds learn new information from a speaker with a history of accuracy rather than from one with a history of inaccuracy (e.g. Birch, Vautier & Bloom, in press; Jaswal & Neely, 2006; Koenig, Clément & Harris, 2004; Koenig & Harris, 2005; Scofield & Behrend, 2008). In a typical task, the child participant observes two speakers name several familiar objects, one of whom consistently gives accurate names, while the other gives inaccurate names for no apparent reason, thereby rendering themselves untrustworthy as informants. The two speakers then offer different names for an unfamiliar object, and the child decides which novel name to accept. Children prefer the name offered by the previously accurate speaker.

What underlies this preference for the previously accurate speaker's information? There is some evidence that with this typical procedure, children simply predicted the accuracy of future output on the basis of previous output, irrespective of the reasons for past inaccuracy and without necessarily making a mentalistic interpretation of the speaker's behavior (Nurmsoo & Robinson, in press *a*). Using a procedure based on that of the authors cited above, Nurmsoo and Robinson found that when the speaker was blindfolded when she was asked to name familiar objects, so her inaccuracy could in principle be excused, children continued to mistrust that speaker when the blindfold was

removed. Such a strategy of generalizing from accuracy of previous output alone, regardless of the reasons for it, reduces the risk of learning falsehoods, but is inflexible and over-cautious, possibly resulting in failure to learn what is true. That is, children may miss out on learning from a currently well-informed and accurate speaker whose past inaccuracy occurred due to circumstances that no longer hold.

A more flexible but also cognitively more demanding strategy, would take into account reasons for past inaccuracy and so allow the possibility of learning from someone whose inaccuracy is excusable due to particular circumstances. Can children use this second strategy? There is evidence to suggest that they can, under some conditions at least. Nurmsoo and Robinson (in press b) used a procedure that differed in several respects from the typical procedure described above (in which children appeared to generalize from accuracy of output alone). Instead of observing two speakers, 3- to 5-year-olds played aimed to identify a hidden toy in collaboration with one puppet. The puppet built up a history of inaccuracy either despite having full information (e.g., he misidentified the color of the target toy despite having seen it), or out of ignorance (he did not have the required information access, for example he misidentified the color of the target toy after having only felt it). On test trials in both conditions the puppet was well informed about the hidden target's identity, while the child could only guess. Children ignored the puppet's well informed suggestion when the puppet had a history of unexplained inaccuracy, but frequently believed him when the puppet's past inaccuracy was due to limited information access. That is, children mistrusted an informant whose previous inaccuracy had no obvious circumstantial explanation, but trusted one whose inaccuracy could be explained in terms of circumstances that no longer held.

In the present study, we explore in two additional ways children's willingness to trust a speaker whose past inaccuracies were due to temporary circumstances. First, we examined whether children were prepared to learn from such a speaker even if his currently well-informed assertion contradicted their well-founded expectation about the target's identity. If they were, this would provide evidence of stronger trust in the speaker than in previous research, in which children had no prior expectation about the true state of affairs.

Second, the inaccurate speaker was not simply ignorant on history trials, but held a false belief. Hence children who understood that people can hold false beliefs (understanding commonly achieved around the age of 4 years e.g. Wellman, Cross and Watson, 2001), could understand not just why the speaker was wrong, but also why he made the specific error he did. Were children who understood this more likely than those who did not, to over-ride their expectation about the target's identity and believe the speaker's contradicting suggestion? In previous research, no relationship has been found between children's false belief understanding and their preference for a previously accurate over a previously inaccurate speaker (Harris, 2007; Pasquini, Corriveau, Koenig & Harris, 2007). However in that research the inaccurate speaker misnamed familiar objects despite apparently having full information, so children had no independent evidence that the inaccuracy could be attributed to a false belief. The interpretation that she was simply untrustworthy might have been more obvious. In the study reported below, in contrast, clear evidence of the speaker's false belief was available.

Method

Participants. Ninety-seven children aged 3;8 to 5;9 (mean 4;7, 55 girls, 42 boys) participated. Five additional children were excluded from the final analysis for providing wild guesses when initially asked the content of the test box (dog, dog, key, cow, nothing), which rendered their second response uninterpretable.

Materials. For the initial unexpected contents false belief test, materials included a juice bottle containing a toy cat, and a small teddy bear. For the puppet familiarization, the hand puppet Dexter and an unmarked box containing a key were used. For the history trials, materials included a box featuring a well-known make of building blocks, a small cereal box, and a DVD case featuring a well-known cartoon character. Each of these history boxes had the atypical content of a small toy cat. For the test trial, a crayon box containing crayons and a small plastic fish was used. Each container was selected to have clear expected contents and pilot testing confirmed that most children did indeed expect the typical contents.

Design and Procedure. Children were randomly assigned to the Ignorant (N = 51) or Unreliable (N = 46) history conditions. Participants were first given an unexpected contents false belief test of the kind widely used in research on children's developing understanding of the mind (e.g. Gopnik & Astington, 1988; Hogrefe, Wimmer & Perner, 1986; Perner, Leekam & Wimmer, 1987; Wimmer & Hartl, 1991). This was followed by familiarization with the puppet, three history trials, a test trial, and a final explanation question designed to ascertain whether children understood that the puppet's false belief was responsible for the particular errors he made on history trials.

The *unexpected contents false belief test* used the juice bottle that unexpectedly contained a toy cat. We followed the typical procedure for an unexpected contents false

belief test: On seeing the unopened bottle children predicted the expected content, were shown the true content, and then were asked both what the teddy bear (absent when the bottle was opened) thinks, and what they themselves had said was inside the bottle before it was opened. Finally, their memory for the true content was checked.

The *puppet familiarization trial* followed, the aim of which was to introduce Dexter and show him to be a reliable informant when he had relevant information access. Children were told that together with Dexter, the game was to identify the content of a box. They were then shown an unmarked box. Dexter reported that he didn't know what was inside, so he would have a look. After looking in the box, without showing the child the contents, Dexter announced that the box contained a key. Dexter was then removed from the scene and the child was asked to make a judgment about what was in the box. The box was then opened, and children discovered that Dexter had made a correct judgment.

The *history trials* followed this familiarization. Children were shown three closed boxes (the box picturing building blocks, the cereal box and the DVD case). For each box in turn, the child was given time to look at the exterior before the Experimenter opened it to reveal the toy cat, which was then returned to the box. In this way, children were led to expect that the final box, the crayons box, also contained a cat. The Experimenter reached for the crayon box and said, "Shall we see what's in here?" before stopping herself and saying, "Actually, before we open this box, let's show this to Dexter, because he hasn't seen these boxes yet." Dexter was then re-introduced. The Experimenter said, "Dexter, we're going to play a game. I'm going to show you some boxes, and you have to say what you think is inside."

For children in the *Ignorant condition*, Dexter looked at each of the boxes in turn, but did not open them. For each box, he announced that the boxes contained the expected contents (blocks, cereal, DVD). In this way, children in the Ignorant condition (who already knew the true contents) witnessed the puppet make three errors while he was poorly informed, and had a false expectation of the contents of the boxes.

For children in the *Unreliable condition*, Dexter opened each of the boxes in turn, peering inside without letting the child see. For each box, he announced inaccurate contents (horse, frog, cow). For example, after looking inside the DVD case (which the child knew contained a cat) Dexter said, “It’s a cow!” In this way, children in the Unreliable condition (who already knew the true contents) witnessed the puppet make three blatant and inexplicable errors despite being well informed, as in the previously cited studies on children’s trust in speakers (e.g. Koenig & Harris, 2005).

The *test trial* followed. In both the unreliable and the ignorant conditions, after Dexter incorrectly reported the contents of the three boxes, the Experimenter said to the child, “He’s not doing very well is he?” before showing Dexter the final crayon box and asking, “What do you think is in here, Dexter?” Dexter asked the child, “I want to get this one right. Can you help me? What do *you* think is in here?” without permitting the child to look. This reminded the children in both conditions that Dexter had been inaccurate, and implied equally in both conditions that Dexter intended to be correct. This also allowed us to check whether or not the child had formed the expectation that the crayon box, like the others, contained a cat.

Dexter looked inside the box, closed it again, and made his judgment, always contradicting the child. Dexter’s judgment varied with the condition, so that he continued

to give the same kind of response that he had on the history trials. Hence for children in the Ignorant condition, Dexter gave an answer consistent with the box's exterior, "crayons," while for children in the Unreliable condition, he gave an answer referring to an unexpected content, "fish." The only exception was for children in the Ignorant condition who predicted that the box contained crayons, having failed to build up the expectation of a cat. For these children, Dexter announced that the box contained a fish. Rather than exclude these children, we analysed their results separately, see below.

We were interested in whether or not children changed their own expectation and believed Dexter's well-informed announcement about the contents of the crayons box. The Experimenter asked the child: "So, when we open up this box, what are we going to find?" The crayons box was then opened, and the Experimenter managed the contents so that Dexter was shown to be correct, to avoid ending the game with a puzzling error.

After a short delay, children were shown one of the history boxes again (e.g., the DVD box), and were asked why Dexter had made an error while reporting its contents.

Results.

Our first hypothesis was that, irrespective of false belief status, children in the Ignorant condition would be more likely to believe Dexter than those in the Unreliable condition: This differentiation could occur so long as children understood "don't see = don't know," understanding achieved by most 3-year-olds (e.g. Pillow, 1989). In the Ignorant condition, Dexter's inaccuracy on the history trials could be excused on the grounds of inadequate information access. In contrast, in the Unreliable condition, Dexter's inaccuracy on history trials could not be excused since he had seen inside the boxes. Children were expected to treat him as an unreliable speaker.

This hypothesis was supported: In the Ignorant condition, 29 out of 51 children (57%) believed Dexter, whereas in the Unreliable condition, 9 out of 46 children (20%) believed him: $\chi^2(1, N = 97) = 12.6, p < .001$. For each expectation sub-group (cat or crayons) examined separately, children in the Ignorant condition were more likely to believe Dexter than those in the Unreliable condition: Fisher exact probability tests, predicted crayons $\chi^2(1, N = 35) = 5.51, p = 0.02$; predicted cat $\chi^2(1, N = 62) = 5.84, p = 0.02$. Of those who predicted cat, 12 out of 27 believed the ignorant puppet, while 6 out of 35 believed the unreliable puppet; of those who predicted crayons, 17 out of 24 believed the ignorant puppet, while 3 out of 11 believed the unreliable puppet.

These results show that children took into account not only the puppet's history of inaccuracy, but also the reasons for it. When inaccuracy could be excused on the basis of inadequate information access, the puppet was more likely to be believed than when there was no obvious circumstantial explanation for the inaccuracy.

Our second hypothesis concerned the Ignorant condition only: That children who understood why Dexter made the particular errors he did on history trials when he held a false belief, would be more likely than those who did not, to believe Dexter on the test trial, when he was well-informed. Although children who only understood ignorance but not false belief had grounds for believing Dexter in the Ignorant condition, we predicted that children who additionally understood why Dexter made the particular errors he did on history trials would be more likely to revise their own well-founded expectation of the contents of the crayons box.

We had two measures of children's false belief understanding: Performance on unexpected contents test, and explanations for Dexter's incorrect answers on the history

trials. For the *unexpected contents test*, we included only children who answered the unexpected contents memory check question correctly, and on these grounds we excluded five nursery children, leaving 46 children in the Ignorant condition for this analysis. Children were classified as false belief failers if they gave no correct answers to either of the two false belief questions (self and other), $N = 21$, Children who gave one or more correct answers were classified as false belief passers, $N = 25$.

For the *explanation question*, all children in the Ignorant condition were categorized into two groups on the basis of their explanation. Children who gave explanations that referred to the pictures on the boxes (e.g., “Saw the picture”; “Because it’s got cornflakes on the front”) were coded as ‘False belief explanation’. All other children, including those who provided no clear response, were coded as ‘No false belief explanation.’ Spontaneous responses in this category included explanations such as “He likes DVDs and he doesn’t like cats” and “Because he got it wrong. He’s a silly duck.” All children’s answers were coded independently by the two authors, and the 5 differences that occurred were resolved by discussion. Twenty three children were classified as giving a false belief explanation and the remaining 27 gave no false belief explanation. There was no significant difference in age between these two groups of children: $t(48) = -1.36, p = .19$.

There was a significant relationship between the two false belief measures: Excluding the 5 children who failed the unexpected contents memory check question, 17 children passed both the explanation and unexpected contents task, 15 failed both, 9 passed only the unexpected contents task, and 5 failed this task but provided a false belief explanation: $\chi^2(1, N = 46) = 7.39, p = .007$.

In line with our hypothesis, the explanations were related to children's decisions to believe or disbelieve Dexter: Eighteen of the 24 (75%) who gave false belief explanations, believed Dexter in the ignorant condition, compared with 11 out of 27 (41%) who gave no such explanations: $\chi^2(1, N = 51) = 6.08, p < 0.05$. However there was no such significant effect for performance on the unexpected contents test: Of the false belief failers, 10 believed and 11 disbelieved the puppet. Of the passers, 14 believed and 11 disbelieved the puppet: $\chi^2(1, N = 46) = .32, p = 0.57$. Perhaps with hindsight it is not surprising that children's ability to infer false belief when asked explicitly in the false belief test did not accurately predict their spontaneous use of false belief reasoning to excuse Dexter's errors on the history trials. Children's explanations in answer to the question about why Dexter made errors provide a more direct measure of children's spontaneous evaluation of Dexter's errors.

As expected, and consistent with previous research (Harris, 2007; Pasquini et al., 2007), in the Unreliable condition there was no relationship between performance on the unexpected contents false belief test and belief or disbelief in Dexter on the test trial.

Discussion and Conclusions

Our main findings are (i) Children were unlikely to believe a currently well-informed speaker who had previously repeatedly been inaccurate for no apparent reason; (ii) they were more likely to believe a currently-well informed speaker whose repeated past inaccuracies could be explained in terms of limited information access; (iii) children believed such an excusable speaker even though he contradicted their well-founded expectation about the true state of affairs; (iv) children who understood not just why such

an excusable speaker had been wrong, but why he had made the particular error he did, were most likely to believe what they were told.

The first finding confirms that of the previous research cited in the introduction (e.g. Birch, Vautier & Bloom, in press; Koenig & Harris, 2005). It may not be surprising that children (and perhaps adults) dismiss assertions from speakers who showed the repeated, implausible inaccuracy found in these studies and in our Unreliable condition. In many of these studies, mistrust on the basis of past inaccuracy is construed by the authors to be appropriate in that it protects children from learning falsehoods. However, although it may be sensible to mistrust a speaker who is blatantly and repeatedly inaccurate for no apparent reason, generalizing on the basis of past accuracy alone can be over-cautious, and could lead children to reject assertions that are likely to be true.

A more flexible approach that takes into account the likely reasons for a speaker's past inaccuracy may increase the chances that children learn what is true and reject what is false. Our second finding listed above indicates that children showed such flexibility. As mentioned in the introduction Nurmsoo and Robinson (in press b) also report such an effect. On the other hand, Nurmsoo and Robinson (in press a) using a procedure based on Koenig and Harris (2005), found no such flexibility, and raise the possibility that features of the communicative interaction between speaker and child are crucial in determining whether children attend only to history of accuracy, or consider possible reasons for speakers' inaccuracy.

The third finding listed above, indicates that many children were prepared to believe the puppet speaker in the Ignorant condition even though they had good reason to expect there to be a cat in the target box on the test trial, having seen cats in the previous

4 boxes. These children apparently assumed that since the puppet had looked inside the target box, his assertion was more likely to be true than their expectation based on experience of the previous boxes, even though the puppet had been inaccurate when he had not looked inside the boxes. This over-riding of children's expectation implies stronger trust in the speaker than was needed in all previous studies in which children had no prior expectation about the truth.

One possibility arising from the fourth result listed above, is that some children were prepared to believe the puppet despite their contradicting expectation, because they could understand not just why he had been inaccurate, but also why he made the specific errors he did. Although the design of our study does not allow us to argue this strongly, children who could explain the puppet's errors in terms of false belief, could have been more confident that the errors occurred due to the circumstances at the time, and that the puppet's longer term trustworthiness was not in doubt.

More broadly, our results are consistent with the possibility that the more children's understanding about the workings of the mind advances, and the more they come to understand about the particular circumstances that can lead speakers to make errors, the more willing they may become to trust and learn from people who have been wrong in the past. They may achieve a balance between the risk of learning falsehoods from people who really are untrustworthy, and the risk of rejecting what is true from people who simply made excusable errors.

References

- Birch, S. A. J., Vauthier, S. A., & Bloom, P. (In press). *Three- and four-year-olds spontaneously use others' past performance to guide their learning*. *Cognition*
- Clément, F., Koenig, M., & Harris (2004). The ontogenesis of trust. *Mind and Language*, *19*, 360-379.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26-37.
- Harris, P. L. (2007). Trust. *Developmental Science*, *10*, 135-138.
- Hogrefe, G., J., Wimmer, H., Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*, 567-582.
- 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., and Harris, P. L. (2004) Trust in testimony: Children's use of true and false statements, *Psychological Science*, *15*, 694–698.
- Nurmsoo, E. & Robinson, E.J. (in press a) *Identifying unreliable informants: Do children excuse past inaccuracy?* *Developmental Science*.
- Nurmsoo, E., & Robinson, E. J. (in press b) *Children's trust in previously inaccurate informants who were well- or poorly- informed: When past errors can be excused*. *Child Development*.

- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216-1226.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125-137.
- Pillow, B. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116-129.
- Scofield, J. & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*, 278-290.
- Wellman, H.M., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.
- Wimmer, H., & Hartl., M. (1991). Against the Cartesian view of mind: Young children's difficulty with own false beliefs. *British Journal of Developmental Psychology, 9*, 125-138.