The Ontogenesis of Trust

FABRICE CLÉMENT, MELISSA KOENIG AND PAUL HARRIS

Abstract: Psychologists have emphasized children's acquisition of information through first-hand observation. However, many beliefs are acquired from others' testimony. In two experiments, most 4-year-olds displayed sceptical trust in testimony. Having heard informants' accurate or inaccurate testimony, they anticipated that informants would continue to display such differential accuracy and they trusted the hitherto reliable informant. Yet they ignored the testimony of the reliable informant if it conflicted with what they themselves had seen. By contrast, three-year-olds were less selective in trusting a reliable informant. Thus, young children check testimony against their own experience and increasingly recognise that some informants are more trustworthy than others.

In recent years, psychologists have described the different processes by which children are able to acquire information about the world. Various knowledge acquisition mechanisms have been proposed and detailed as 'naïve theories', enabling infants and young children to get a grip on naive physics (Baillargeon, 1993; Spelke, 1991), arithmetic (Carey, 1988; Gelman and Gallistel, 1978; Wynn, 1992), psychology (Astington, Harris and Olson, 1988; Wellman 1990), biology (Medin and Atran, 1998; but see Carey, 1995) and even sociology (Hirschfeld, 1996; Kaufmann and Clément, 2003).

The general impression created by this literature is that young children elaborate such fundamental conceptual knowledge by themselves, without outside assistance. However, if we could invent a machine able to compute the number of beliefs held by any individual, we would find that an extremely large number of them were not discovered in a 'social vacuum'. On the contrary, the child's developing stock of beliefs is considerably enriched by information communicated by *testimony*. These beliefs can be grossly divided into three categories. First, children presumably learn about many past events via testimony. For example, they will learn about past episodes in the lives of their parents by listening to family

The first author was supported by a grant from the Swiss National Fund (8210-061282). Study 1 was carried out at the Institute for Human Development, University of California, Berkeley. We are grateful to members of this department for their advice and help, especially Alison Gopnik, David Sobel, Tamar Kushnir, Jane Perry, Joe Campos and Diana Baumrind. Study 2 was carried out when visiting the Culture and Cognition program, University of Michigan, Ann Arbor. We thank the staff of the Institute for Social Research, especially Laura Reynolds, and the following faculty members for their invaluable help: Lawrence Hirschfeld, Henry Wellman, Susan Gelman and Richard Nisbett. Melissa Koenig was supported by NICHD Grant 1 F32 HD42860. At Harvard, we received helpful comments from Rita Astuti and Laurence Kaufmann.

Address for correspondence: Dr. Paul L. Harris, 503A Larsen Hall, Harvard Graduate School of Education, Appian Way, Cambridge, MA, 02138, USA.
Email: harrispa@gse.harvard.edu

Mind & Language, Vol. 19 No. 4 September 2004, pp. 360–379.

narratives. Similarly, they will learn about key historical events from the testimony of others. Indeed, it is plausible to suppose that children's cumulative sense of the way in which their own personal biography is situated within a larger, historical narrative depends crucially on the testimony that is provided by other people. Second, children presumably learn about many scientific facts or claims via testimony insofar as they lack the relevant observational capacities. For example, children are not likely to discover by themselves the spherical shape of the Earth or the relationship between the mind and the brain (Harris, 2002a). Third, testimony is essential for the acquisition of beliefs about metaphysical powers that are not accessible to any empirical observation. For example, it has been shown that children are ready to accept the existence of supernatural beings endowed with extraordinary capacities, like omniscience (Barrett, 2000) or eternal life (Giménez and Harris, 2001). History, science and religion are, therefore, three domains where children's own cognitive efforts cannot be the unique source of their knowledge: to acquire these kinds of representations, children rely on others' testimony.

If it is admitted that children are not 'stubborn autodidacts' and frequently have to rely on communicated information, we face an important issue: to what extent are children ready to believe what they are told by others and in particular by adults? From a philosophical point of view, this question relates back to an important epistemological question. How can we characterize the learning processes by which information that is not acquired directly via first-hand observation is nevertheless held to be true? In other words, how is the indirect acquisition of knowledge to be justified? We have claimed that a strict empiricist perspective, which gives absolute priority to personal and first-hand experience, cannot account for the acquisition of scientific and metaphysical beliefs. Children, and no doubt adults, must often resort to something different from personal experience. To resolve this epistemological problem, Thomas Reid (1785) proposed that perception and testimony are actually comparable, testimony simply being a form of indirect perception. More recently, Ruth Millikan has defended a similar position in claiming that language, like light, is a direct medium of perception (Millikan 1998, p. 64). If children adopt this position of trust toward testimony, the transmitted propositions would automatically induce certain belief states and the receiver would be prepared to act as though the proposition were true (Gilbert, 1991; 1993). It is only in a subsequent phase that the representation might potentially be called into question and revised. We propose to call this position 'indiscriminate trust'.

A positive feature of this conception is that it gives an account of the long 'chain of trust' that has characterized—and continues to characterize—the transmission of knowledge in the human species. However, it ignores a crucial element of many social environments in which communication takes place: informants can deliberately transmit erroneous information for their own advantage. According to evolutionary psychologists, the risk of deception was so high that it led to an escalating arms race where competitors were eventually forced to develop

mind-reading mechanisms to decipher others' intentions (Krebs and Dawkins, 1984). If we hypothesize, in line with evolutionary psychology, that this competitive context played a decisive role in the shaping of our psychological mechanisms (Cosmides and Tooby, 1989), it would be surprising if all communicated information were, by default, considered as true by the cognitive system. As Perner put it: '. . . (it) would be disastrous if (the knowledge base) were linked (causally) to symbolic input, like language. Because linguistic statements can be unreliable (mistakes, lies) and are subject to frequent errors of interpretation, the child's knowledge base would become alarmingly unstable'. (Perner, 1988: p. 145). A form of automatic caution should, on the contrary, submit the transmitted proposition to a basic checking procedure, in order to test its validity (Sperber, 2001).

On this argument, we would expect children to recognize and reject statements which conflict with information that they have previously acquired via first-hand observation. In line with this expectation, there is evidence that infants and young children contradict and seek to correct informants who assert names for objects that are inconsistent with the children's own past experience (Koenig and Echols, 2003; Pea, 1982). Similarly, Robinson, Mitchell and Nye (1995) report that preschoolers quite often reject an adult's claim about the contents of a toy box if that claim is inconsistent with the apparent contents illustrated on the outside of the toy box. Moreover, not only do preschoolers resist empirically dubious claims themselves, they realize that other people will give priority to what they have seen over what they are told. For example, they judge that someone who has seen orange juice in a jug will likely ignore a contradictory claim that there is milk in the jug (Mitchell, Robinson, Nye and Isaacs, 1997).

Despite convergent evidence that young children check an informant's claims against their own first-hand experience, we do not know if children keep track of those checks in order to assess whether particular informants have proven more or less reliable. In principle, children could conclude that certain informants are reliable insofar as their claims regularly coincide with first-hand observation whereas others are unreliable because their claims regularly conflict with first-hand observation. Subsequently, in the absence of relevant first-hand information, children might be prepared to trust testimony if it is provided by someone who has proved reliable hitherto but not if it is provided by someone who has proved unreliable hitherto. We shall refer to such cautious and selective acceptance of testimony as 'sceptical trust'.

Finally, we may ask how young children respond when an informant who has hitherto provided reliable information suddenly says something that contradicts what children have observed for themselves. Children who display sceptical trust, in the sense defined above, might respond in two different ways. First, if children give ultimate priority to first-hand experience—as implied by the sceptical position—they should resist the informant's testimony despite his or her past reliability. Alternatively, if children suspend or inhibit their normally routine, empirical checking—having come to trust a consistently reliable informant— they might

accept testimony from that informant even if it contradicts what they have observed for themselves. We shall refer to this position as 'gullible trust'.

Experiment 1 was designed to explore what stance children take toward others' testimony. Is their trust in testimony indiscriminate, sceptical, or gullible? To examine this question, children were first presented with two informants: one who described three objects accurately and one who described the same three objects inaccurately. Children were then tested on a series of tasks involving the two informants. In the 'prediction' task, children's grasp of the difference between these two informants was checked by asking them to say how each informant would describe a fourth object. In the 'convergence' task, children's accuracy at describing an object was assessed when first-hand observation and testimony from a reliable informant were in agreement. The 'guessing' task was more challenging: it was designed to assess whether children, when lacking perceptual access, would display indiscriminate trust or sceptical trust in testimony. The reliable informant described a hidden object in one way and the unreliable informant described it in a different way. Children were then invited to provide their own description. The experimental question was whether they would display indiscriminate trust by agreeing at random with one of the two informants or sceptical trust by agreeing with the hitherto reliable informant. The contradiction task was designed to assess whether children would display sceptical or gullible trust when the hitherto reliable informant contradicted their first-hand experience. Children were briefly shown an object which was then hidden. Not just the unreliable informant but also the hitherto reliable informant offered a description of the object that conflicted with what children had seen for themselves. Children were then invited to provide their own description. The experimental question was whether children would display gullible trust by continuing to agree with the hitherto reliable informant or sceptical trust by giving ultimate priority to what they themselves had seen.

1. Experiment 1

1.1 Method

1.1.1 Participants. Twenty-six younger children, most of them 3-year-olds (Range = 36 to 50 months; M = 3 years 9 months) and twenty-seven older children, all of them 4- and 5-year-olds (Range = 51 to 70 months; M = 4 years 10 months) participated. Children were interviewed in Berkeley, CA, at the University of California childcare centre. The majority came from middle and upper-middle class families. They were interviewed individually, in a small room, for about 10 minutes.

1.1.2 Materials. Two puppets were introduced as informants: a frog ('Froggie') and a mouse ('Mousie'). Various objects were presented to the children: some fruit (an apple or a banana), a teddy bear, and pompons of different colours. When

not in use, the pompons were all placed in a small black bag. During the interview, individual pompons were taken from the bag and put inside a rectangular, cardboard box.

1.1.3 Procedure. The experimenter and the child sat on opposite sides of a little table. After being seated, the experimenter introduced the puppets. One puppet was described as very nice. The experimenter also explained that: 'Each time you ask him a question, he gives you the right answer'. The other puppet was also characterized as very nice, but the experimenter explained that: 'Each time you ask him a question, he gives you some strange answers'. For approximately half of the children, the frog puppet served as the reliable source, and for the other half, the mouse puppet served as the reliable source.

Next, the experimenter proposed that they see what would happen when different objects were put on the table. A first object (either an apple or banana) was presented to the child and to the puppets. The experimenter asked both puppets to tell them what was on the table. The reliable puppet correctly stated what was on the table (e.g., 'a banana'), whereas the unreliable puppet said something incorrect (e.g., 'an orange'). A second object (a teddy bear) was presented. The experimenter asked the puppets what it was and again the reliable puppet identified it correctly whereas the unreliable puppet did not. Finally, one of the pompons was taken from the bag and put on the table. The experimenter asked the puppets about its colour. The reliable puppet stated the correct colour, whereas the unreliable source puppet stated a different colour. In sum, children were presented with one informant's accurate testimony and the other informant's inaccurate testimony over three successive occasions.

1.1.4 Prediction Task. The experimenter took another pompon from the bag and put it on the table. He then asked first with respect to the reliable puppet and then with respect to the unreliable puppet: 'Now, if I ask Froggie/Mousie about the colour of the pompon, what do you think Froggie/Mousie will say?' Children who stated the visible colour for both puppets were given an additional prompt. The experimenter put another pompon on the table, said: 'Let's try with another one . . . remember that Froggie/Mousie says strange things' and asked the test question again with respect to each puppet. Children who failed to answer correctly with respect to each puppet (by stating the visible colour for the reliable puppet and another colour for the unreliable puppet) were not tested any further.

1.1.5 Convergent Task. The experimenter showed the child a cardboard box, said that it was a 'pompon box', and explained that he would put some pompons into the box. He took one of the pompons from the black bag, put it on top of the box for about three seconds, and then put it into the box, so that it remained invisible for the remainder of the task. Next, the experimenter made the two puppets 'look' into the box, and asked each of them about the colour of the

pompon inside. First, the reliable puppet correctly stated the colour that was in the box, and then the unreliable puppet stated a different colour. Finally, children were asked: 'And you, can you tell me what is the colour of the pompon in the box?'

1.1.6 Guessing Task. After removing the pompon from the box without showing it to the child, the experimenter took another pompon from the bag, explained to the child that he or she was not going to see the pompon this time and put it into the box without letting the child see its colour. The experimenter again made the two puppets 'look' into the box, and asked each of them about the colour of the pompon inside. First, the reliable puppet correctly stated the colour that was in the box, and then the unreliable puppet stated a different colour. Finally, children were asked: 'And you, can you guess what is the colour of the pompon in the box?'

1.1.7 Contradiction Task. The pompon was again removed from the box in such a way that the child could not see it. Another pompon was taken from the black bag and the bag was set aside (either on the table or on the ground, next to the experimenter). The pompon was put on top of the box for about three seconds, and then hidden in the box. As before, the two puppets were made to 'look' into the box and asked about the colour of the pompon inside. For this task, both the unreliable puppet followed by the reliable puppet stated a different colour from the actual colour of the pompon in the box. Children were then asked: 'And you, can you tell me what is the colour of the pompon in the box?' Finally, children were asked to justify their response: 'How do you know that the pompon is x?',—x being the colour stated by the child.

1.2 Results

1.2.1 Prediction Task. Recall that the children were asked to predict what the reliable and unreliable puppets would say about the colour of a pompon fully visible on the table. Table 1 shows the proportion of children in each age group who correctly stated the actual colour of the pompon when asked about the reliable puppet and appropriately stated a different colour when asked about the unreliable puppet. Inspection of Table 1 shows that children in both age groups were quite accurate at predicting what the reliable puppet would say. With only

	Type of Puppet	
Age	Reliable	Unreliable
Younger	0.96	0.58
Older	1.00	0.85

Table 1 *Proportion of children making correct predictions as a function of age and type of puppet.*

one exception (in the younger group), they all answered that the reliable puppet would identify the colour of the pompon correctly. Children were less accurate in stating what the unreliable puppet would say, particularly in the younger group. Children erred by predicting that the unreliable puppet would also say the correct colour. Nevertheless, McNemar tests for the significance of changes confirmed that children in each age group typically differentiated between the two puppets: they were more likely to state the visible colour for the reliable puppet but not for the unreliable puppet than to do the reverse both in the younger group (χ^2 (N = 14) = 9.60, $p < .01$) and in the older group (χ^2 (N = 23) = 21.04, $p < .001$).

In the younger group, 14 out of 26 children (54%) were correct for both puppets and in the older group, 23 out of 27 children (85%) were correct for both puppets. A Chi-square test showed that the proportion of children who were correct for both puppets was greater in the older group than in the younger group (χ^2 (N = 53) = 13.77, $p < .01$.

1.2.2 Convergent Task. Recall that only children who passed the prediction task by responding that the reliable puppet would say the visible colour and the unreliable puppet would say a different one proceeded to the convergent task (younger children, N = 14; older children, N = 23). In the convergent task, children were asked to state the colour of a pompon that they had seen originally and which the reliable puppet had subsequently named correctly and the unreliable puppet had named incorrectly. As expected, children in both age groups were relatively accurate (younger group, percentage correct = 78.6%; older group, percentage correct = 95.7%). Binomial tests confirmed that this percentage approached significance for the younger children ($p < .059$) and was greater than would be expected by chance for the older children ($p < .002$). Although the proportion of children who were correct was somewhat greater in the older as compared to the younger group, a Chi-square test fell short of significance (χ^2 (N = 43) = 0.72, n.s.).

1.2.3 Guessing Task. All children who had received the convergent task proceeded next to the guessing task (younger children, N = 14; older children, N = 23). For this task, children were asked to state the colour of a pompon that they themselves had not seen but which the two puppets had seen and named. In the younger group, 71.4% gave the same answer as the reliable puppet; in the older group, 87.0% gave the same answer as the reliable puppet. Binomial tests showed that this percentage was not significantly different from chance for the younger group ($p < .18$) but was greater than would be expected by chance for the older group ($p < .002$). Nevertheless, a Chi-square test failed to confirm that older children were more likely than younger children to concur with the reliable puppet (χ^2 (N = 37) = 0.54, n.s.).

1.2.4 Contradiction Task. With the exception of one older child who was unwilling to continue, all children who had received the guessing task proceeded

next to the contradiction task (younger children, N = 14; older children, N = 22). Children were asked to state the colour of a pompon that they had seen earlier but for this task, both the reliable and the unreliable puppets named the colour of the pompon incorrectly. In the younger group, 64.3% correctly stated the colour that they had seen, 14.3% agreed with the reliable puppet, none agreed with the unreliable puppet, and 21.4% suggested a different colour altogether. In the older group, 68.2% correctly stated the colour that they had seen, 27.3% agreed with the reliable puppet, 4.5% agreed with the unreliable puppet and none suggested a different colour. Thus, the most frequent response in both age groups was to correctly re-state the colour that they had seen.

To check whether the proportion of correct replies exceeded chance, we made the assumption that children behaving randomly would have been correct on 1/3 of all trials in that they were offered a choice among three alternatives (the correct colour; the colour stated by the reliable puppet; and the colour stated by the unreliable puppet). Admittedly, children could, in addition, generate their own alternatives but we chose to assume—conservatively—that children chose only among three options. A binomial test confirmed that in the younger group, the number of correct responses was greater than chance (N = 14, $p < .02$). In the older group, the number of correct replies also exceeded chance (χ^2 (N = 22) = 12.02, $p < .001$). The proportion of children who ignored the evidence of their own eyes and went along with the reliable puppet was slightly greater in the older group than the younger group (27.3% versus 14.3%) but this age change fell well short of significance (Fisher Exact Probability Test (N = 36), $p = 0.31$).

Following their choice of colour, children were asked: 'How do you know that it's ____?' Children's replies were allocated to three different categories: *Saw* (e.g. 'Because I saw it' or 'I saw you put it there'); *Reliable Puppet* (e.g. 'Because he (i.e. the reliable puppet) said it'*; Uninformative* (e.g., 'Don't know' or 'Because!' or a non response). Collapsing across the two age groups, a total of 19 children (3 younger children and 16 older children) provided informative justifications (i.e., either *Reliable Puppet* or *Saw* justifications). Among these children, 14 had stated the colour that they had seen whereas the remaining 5 had agreed with the reliable puppet. These two sub-groups differed in the pattern of their justifications. All 14 of the former group offered *Saw* justifications. By contrast, only one child in the latter group offered a *saw* justification and the other four mentioned the *Reliable Puppet*. A Fisher Exact Probability Test confirmed that this difference between the two subgroups in their pattern of justifications was significant ($p < .01$).

1.3 Discussion. The older children in Experiment 1 responded in a very systematic fashion. First, most of them passed the prediction task by correctly anticipating what the reliable and unreliable puppets would say. In the convergent task, where first-hand observation and the testimony of the reliable puppet coincided, they correctly specified the colour of the pompon. In the guessing task, where only the conflicting testimony of the two puppets was available, they agreed with the reliable puppet. Finally, when their own observation and the testimony of

the reliable puppet conflicted, most older children relied on their own observation—and when they provided a justification, such children referred back to what they had seen. In summary, older children adopted a position of sceptical trust. This conclusion finds support in two main findings: Older children differentiated between the two informants, and in the absence of first-hand observation, they trusted the testimony of the hitherto reliable informant. Nevertheless, when first-hand information conflicted with the testimony of the hitherto reliable informant, they gave priority to their own first-hand observation. They showed little sign of gullible trust in the reliable informant.

Some of the younger children also adopted the position of sceptical trust but overall the response pattern of the younger children was less systematic. Nearly half of the younger children failed the prediction task, typically by saying that both puppets would state the actual colour of the pompon. The majority of the younger children who passed the prediction task went on to pass the convergence task and to agree with the reliable puppet in the guessing task but neither of these trends reached significance. Their response pattern in the contradiction task was more clear-cut. Consistent with the position of sceptical trust and consistent with the pattern displayed by most of the older children, they gave priority to their own first-hand observation rather than the testimony of the reliable informant.

Experiment 2 was designed to check on these initial findings with a new sample of children and also to assess potential age changes more thoroughly. To this end, various changes were introduced in the procedure. First, it will be recalled that children who failed the prediction task did not proceed to subsequent tasks in Experiment 1. To the extent that a large number of children in the younger group failed the prediction task and were eliminated from further testing, the resulting sample of younger children may have been unrepresentative and thereby attenuated potential age differences. Accordingly, in Experiment 2, children proceeded to later tasks irrespective of whether or not they had passed the prediction task. Second, various changes in word order were introduced. Third, following children's replies to the contradiction task, they were given a search task; this provided a non-verbal index of what colour they attributed to the pompom. These changes are described in detail in the method section.

2. Experiment 2

2.1 Method

2.1.1 Participants. Twenty-two younger children, all of them 3-year-olds (Range = 36 to 45 months; M = 3 years 7 months) and twenty-eight older children, most of them 4-year-olds (Range = 46 to 60 months; M = 4 years 4 months) participated. Children were interviewed in Ann Arbor, MI, in two University of Michigan childcare centres. The majority came from middle and upper-middle class families. Children were interviewed individually, in a small room, for about

10 minutes. When parents gave authorization, the sessions were videotaped (N = 33).

2.1.2 Materials and Procedure. The materials and procedure for Experiment 2 were the same as those for Experiment 1 with the following changes. First, children continued with the testing procedure irrespective of whether they passed or failed the prediction task. Second, the order of mention of the reliable versus the unreliable puppet during questioning was systematically varied within and across participants. Finally, a search task was added at the end of the procedure as described below.

2.1.3 Search Task. Following the contradiction task, a follow-up task was added. With the pompon still in the box, the experimenter said that the game was over and that the child had done very well. He then asked the child to help to put things away for the next child and asked: 'Can you give me the *x* pompon, please?'—where *x* stands for the colour of the pompon that was actually in the box (and which the child had seen). Thus, children could choose to either remove the pompon from the box where they had seen it placed or look for it in the nearby black bag, where all the other pompons were located. This task offered a non-verbal index of what colour children ascribed to the pompon in the box.

2.2 Results

2.2.1 Prediction Task. Recall that children were asked to predict what the reliable and unreliable puppets would say about the colour of a pompon fully visible on the table. Table 2 shows the proportion of children in each age group who correctly stated the colour of the pompon when asked about the reliable puppet and appropriately stated a different colour when asked about the unreliable puppet. Inspection of Table 2 shows that children in both age groups were very accurate at predicting what the reliable puppet would say. With only one exception (in the younger group), they all predicted that the reliable puppet would identify the colour of the pompon correctly. Children were less accurate in stating what the unreliable puppet would say, particularly in the younger group. In both age groups, children erred by predicting that the unreliable puppet would say the

	Type of Puppet	
Age	Reliable	Unreliable
Younger	0.91	0.55
Older	1.00	0.86

Table 2 *Proportion of children making correct predictions as a function of age and type of puppet.*

correct colour. Nevertheless, McNemar tests for the significance of changes confirmed that children in both age groups typically differentiated between the two puppets: they were more likely to state the visible colour for the reliable puppet but not for the unreliable puppet than to do the reverse both in the younger group (χ^2 (N = 13) = 4.92, $p < .05$) and in the older group (χ^2 (N = 24) = 22.04, $p < .001$).

In the younger group, 11 out of 22 children (50%) were correct for both puppets and in the older group, 24 out of 28 children (86%) were correct for both puppets. A Chi-square test confirmed that the proportion of children who were correct for both puppets was greater in the older group than the younger group (χ^2 (N = 50) = 16.88, $p < .001$).

2.2.2 Convergent Task. With the exception of four younger children, who were unwilling to continue, all children were presented with the convergent task (younger children, N = 18; older children, N = 28). Children were asked to state the colour of a pompon that they had seen earlier and which the reliable puppet named correctly and the unreliable puppet named incorrectly. Children in both age groups were quite accurate (younger group, percentage correct = 88.9%; older group, percentage correct = 89.3%). Binomial tests confirmed that this percentage was greater than would be expected by chance for both age groups ($p < .001$). A Chi-square test confirmed the proportion of children who were correct did not vary between the two groups (χ^2 (N = 46) = 0.13, n.s.).

2.2.3 Guessing Task. With the exception of one older child who was unwilling to continue, all children who had received the convergent task proceeded to the guessing task (Younger children, N = 18; Older children, N = 27). For this task, children were asked to state the colour of a pompon that they had not seen earlier but which the two puppets had seen and differentially named. In the younger group, 50% gave the same answer as the reliable puppet; in the older group, 85.2% gave the same answer as the reliable puppet. Binomial tests showed that this percentage was no greater than would be expected by chance for the younger group but significantly above chance for the older group ($p < .001$). A Chi-square test confirmed that the older children were more likely than the younger children to concur with the reliable puppet (χ^2 (N = 45) = 8.33, $p < .01$).

2.2.4 Contradiction Task. All children who received the guessing task proceeded next to the contradiction task (N = 18 for the younger group; N = 27 for the older group). For this task, children were asked to state the colour of a pompon that they had seen earlier but which both the reliable and the unreliable puppets had named incorrectly. In the younger group, 72% correctly stated the colour that they seen, 11% agreed with the reliable puppet, none agreed with the unreliable puppet and 16.7% either suggested a different colour altogether or said that they did not know. In the older group, 66% correctly stated the colour that they had seen, 22% agreed with the reliable puppet, 3.7% agreed with the unreliable puppet

and 7.4% either suggested a different colour altogether or said that they did not know. Thus, the most frequent response in both age groups was to correctly re-state the colour that they had seen. To check whether the proportion of correct replies exceeded chance, we again made the assumption that children behaving randomly would have been correct on one third of all trials in that they faced a choice among three alternatives (the correct colour; the colour stated by the reliable puppet; and the colour stated by the unreliable puppet.) The number of correct replies exceeded chance for both the younger group (χ^2 (N = 18) = 12.25, $p < .001$) and the older group (χ^2 (N = 27) = 13.50, $p < .001$). The proportion of children who ignored the evidence of their own eyes and went along with the reliable puppet was slightly greater in the older group than the younger group (22% versus 11%). However, this age change fell well short of significance (Fisher Exact Probability Test (N = 36), $p = 0.31$).

Following their answer to the contradiction test question, children were asked: 'How do you know that it's ____?' Children's replies were again allocated to 3 different categories: *Saw*; *Reliable Puppet*; and *Uninformative*. Collapsing across the two age groups, a total of 20 children (7 younger children and 13 older children) provided informative justifications (i.e., either *Reliable Puppet* or *Saw* justifications). Among these children, 17 had correctly stated the colour that they had seen whereas the remaining 3 had agreed with the reliable puppet. These two subgroups differed in the pattern of their justifications. All but one of the former group offered *Saw* justifications. By contrast, no child in the latter group offered a *Saw* justification—all three mentioned the *Reliable Puppet*. A Fisher Exact Probability Test confirmed that this difference between the subgroups in their pattern of justifications was significant ($p < .01$).

2.2.5 Search Task. Recall that the interviewer completed the test procedure of Experiment 2 by asking children: 'Can you give me the *x* pompon, please?'—where *x* stands for the colour of the pompon that was actually in the box. A total of 30 children actively responded to the interviewer's request by searching either in the box or elsewhere (e.g. in the bag holding the pompons, under the table, etc.). Among these 30 children, 22 had correctly stated the colour that they had seen whereas the remaining 8 had agreed with the reliable puppet. These two subgroups differed slightly in their response pattern. All children in the former group looked in the box. By contrast, 6 of the children in the latter group looked in the box and the remaining 2 children searched elsewhere. A Fisher Exact Probability Test showed that this difference between the subgroups approached significance $p < .064$).

2.3 Discussion

The older children in Experiment 2 again responded systematically, just as they had done in Experiment 1. Almost all of them passed the prediction task by correctly anticipating what the reliable and unreliable puppets would say. In the convergent

task, where the testimony of the reliable puppet coincided with their own prior observation, older children correctly specified the colour of the pompon. In the guessing task, where they had not seen the pompon, they agreed with the reliable puppet. Finally, when their own observation and the testimony of the reliable puppet were in conflict, the majority of the older children relied on their own observation and—as in Experiment 1—such children explicitly justified their response by reference to what they had seen. In summary, the results of Experiment 2—which involved a new sample and various procedural changes—confirmed the conclusion reached following Experiment 1. Four-year-olds adopt a position of sceptical trust. They are able to identify who is a reliable informant and in the absence of first-hand observation, they trust the testimony of such an informant. However, when first-hand observation conflicts with testimony—even from an informant who has proved reliable in the past—they do not display gullible trust in that informant. Instead, they privilege their own observation.

In the younger group, a considerable number of children failed the prediction task and—as in Experiment 1—this was often because they expected both puppets to name the colour accurately. When younger children had the opportunity to observe the pompon—as they had in the convergence and contradiction task, they named its colour accurately. By contrast, when they had not had the opportunity to observe the pompon, they responded with indiscriminate trust: they were as likely to agree with the unreliable as the reliable puppet. Thus, younger children resembled older children in displaying scepticism toward testimony when it conflicted with their own first-hand observation. They differed from the older children by displaying indiscriminate trust in testimony in the absence of first-hand observation.

Recall that there was an important difference in the procedure of Experiment 2 as compared to Experiment 1. Even if children failed the prediction task, they were tested on subsequent tasks. In practice, this meant that the composition of the older group changed very little between the two experiments whereas the younger group in Experiment 2 included several children who did not differentiate between the two informants. We may conclude, then, that younger children's display of indiscriminate trust in Experiment 2—their willingness to agree with the reliable and unreliable informants more or less equally—was probably due to their initial difficulties in differentiating between the two informants in terms of their reliability. In particular, younger children were often unable to anticipate that an informant might offer misleading or inaccurate testimony.

3. General Discussion

In this paper, our general goal was to shed light on the extent to which children trust other people's testimony. More specifically, the objective was to explore how three- to five-year-old children use testimony provided by two different informants, one reliable and the other not, in their belief formation process. Previous

research has shown that on any given occasion, young children assess an informant's testimony against their own experience. We asked whether children keep track of an informant's record of reliability and use that record to assess his or her future reliability—particularly in situations where their own experience offers no guidance. The theoretical debate at stake opposes a position of indiscriminate trust, in which children are inclined to accept any communicated information (at least, initially), to a position of sceptical trust, which postulates on the contrary a kind of suspicion or caution. The findings from the older children, and to a less systematic extent those from the younger children as well, indicate that children are indeed cautious. Not only do they check what other people say against what they know to be the case they also use this checking procedure to discriminate between reliable and unreliable informants and to trust the former rather than the latter. Nevertheless, their trust does not evolve into gullibility. If a hitherto reliable and trusted informant says something that conflicts with children's prior knowledge, they reject it. Thus, children displayed what we have referred to as sceptical rather than gullible trust. Indeed, if we examine the pattern of responding displayed by individual children in study 2, we see that 59.3% of older children and 33.3% of younger children conformed to the position of sceptical trust by responding accurately and appropriately on all four main tasks (prediction, convergence, guessing and contradiction).

However, this general conclusion ignores two important age changes. We now turn to these age changes, starting with the prediction task. In the prediction task, children were asked to predict what the two puppets would say when presented with a coloured pompon. Almost all children responded accurately when asked to predict what the reliable puppet would say about its colour. However, an interesting age effect came to light when children had to predict what the unreliable puppet would say. Insofar as the unreliable puppet had systematically given wrong answers in the past, the 'logical' answer was to say that the puppet would name a different colour from the visible one. Older children did precisely this but younger children clearly had difficulty in answering this question. A considerable proportion of them (42% in Experiment 1 and 41% in Experiment 2) maintained that the unreliable puppet would say the *actual* colour of the pompon. An age change also emerged in the guessing task. Given the differential accuracy of the two informants in the familiarization period, it was appropriate to trust only the more reliable informant in the guessing task. Older children displayed this pattern of selective trust. Younger children, by contrast, were indiscriminate: they were just as likely to agree with the unreliable informant as the reliable informant. Thus, even though younger children could be sceptical of an informant's testimony if it conflicted with their own experience (as shown by their replies in the contradiction task), they were not necessarily sceptical of an informant's current testimony if on some earlier occasions his or her testimony had conflicted with their own prior experience. Taken together, these two age changes suggest that there is an important shift between 3 and 4 years of age in children's trust in testimony: older but not younger children recognize that even when they themselves have no experience against

which to check the current testimony of a particular informant, that testimony may be wrong. How can we explain this age change?

Two kinds of explanation may be considered: one focuses on communication, the other on mental states. Three-year-olds may have had recourse to a well-established principle of communication, namely that there is a regular and reliable connection between what people say and the objects they refer to. This connection between speakers and accurate labelling seems to be established early: 16-month-old infants look surprised when a person labels objects incorrectly and attempt to correct the speaker (Koenig and Echols, 2003). This expectation will lead to the following general formula: when someone is asked to name or refer to an object, they will produce the name of the object in question. Guided by this generic formula, 3-year-olds expect both puppets to be accurate and, as a result, err in both the prediction task and the guessing task. Four-year-olds, on the other hand, appear to keep track of, and generalize from the previous speech acts of a given informant. They therefore predict that the unreliable puppet will continue to label the object inaccurately and they eschew his testimony in the guessing task.

To introduce the other possible explanation, we note that the majority of children displaying difficulty with the prediction task were under four years of age. This threshold suggests that their errors could be related to a difficulty in the processing of mental states. The way in which the unreliable puppet acts is not easy to conceptualize because it deviates from a normal communicative act. Usually, as we have noted, people label things accurately. Thus, communication typically involves the transmission of true beliefs. In the prediction and guessing task, however, the children's task is far from obvious because they have to recognize that communication may involve the transmission of false beliefs. The ability to conceptualize false beliefs emerges between 3 and 4 years (Perner, 1991; Sperber, 2000; Wellman, Cross and Watson, 2001; Whiten and Byrne 1991). We can therefore suggest that older but not younger children respond appropriately in the prediction and guessing tasks because they recognize that one of the informants entertains—or seeks to create—a false belief.

One way to decide between these competing explanations is to study the scope of children's selective trust. If the focus on children's developing sensitivity to an informant's communication history is correct, we would not expect the same pattern of selective trust to emerge when children observe adults engaging in non-communicative as opposed to communicative acts. By contrast, if the focus on children's developing sensitivity to false beliefs is correct, we would expect the same pattern, whether adults engage in non-communicative or communicative acts, so long as those acts could involve false beliefs. Thus, children might observe two adults engaged in a sequence of non-communicative acts; for example, one adult uses a series of objects correctly and one adult uses them incorrectly. On the false belief hypothesis but not on the communication hypothesis, older children should be sceptical of information provided by the incorrect model even when he or she engages in such non-communicative acts.

Whatever the exact interpretation of the age change observed in Experiments 1 and 2, it is interesting to compare the current findings with those that have emerged in a series of studies by Robinson and her colleagues (Robinson Champion, and Mitchell, 1999; Robinson and Whitcombe, 2003). Children experienced a contradiction between what they claimed to be inside a box and the claim made by an adult. Children generally resolved this contradiction appropriately by weighing the bases for the two conflicting claims. For example, they were likely to revise their initial claim in favour of the adult's if they knew that the adult had looked inside the box whereas they themselves had simply made a guess. Conversely, they were likely to maintain their initial claim if the adult had not looked inside the box whereas they themselves had done so. Unlike the current set of results, no age change was found. Thus, children ranging from 3- to 6-years displayed the same pattern of sensitivity to the better-founded claim. Moreover, children displayed that sensitivity even though they often made errors in explicitly reporting who had had access to what information. Taken together with the present findings, these results reinforce the conclusion that children display key features of sceptical trust from an early age. In particular, they weigh testimony against empirical observation even if they continue to refine their monitoring and appraisal of informants in the course of the preschool years.

Further evidence for the early weighing of testimony against empirical observation emerged in the contradiction task. This task was designed to see if children would shift from sceptical to gullible trust. More specifically, once children are accustomed to trusting an apparently reliable source, are they ready to believe information that contradicts something they already know? To explore this issue, perception and testimony were put into competition. Children saw the colour of the pompon when the experimenter put it into the box. Then, both the reliable and the unreliable puppets stated the wrong colour. The experimental question was whether children would 'follow' the reliable puppet, even if the information communicated was wrong. The majority of younger children responded with the colour they had just seen but, as just discussed, younger children had not systematically trusted the reliable puppet. Hence, the results from the older group are more informative because the majority of these children had been able to discriminate between the two puppets and to use the reliable informant. It would not be surprising if they were prone to gullible trust by continuing to favour the reliable informant in the contradiction task. After all, this informant had been right in the past. However, this theoretical outcome was strongly contradicted by the fact that most of the older children went with their perception and rejected the communicated information. Moreover, children were able to remember how they obtained this information when asked for a justification. By implication, even when children display selective trust in reliable informants, they do not cease to deploy an empirically-grounded, filtering device. They continue to check what an informant says against what they have observed for themselves, arguably in a relatively automatic fashion. When a contradiction is detected, the communicated statement is rejected in favour of knowledge acquired in a more direct way.

Admittedly, it could be argued that the strength of the results in the contradiction task is due to the nature of the stimuli. After all, how could a child be misled by such an obvious contradiction between what they had seen a few seconds earlier and what the reliable informant claimed? To respond to this criticism, we note that a certain number of children actually did give the same answer as the reliable informant (27% of the older children in Experiment 1 and 22% in Experiment 2). Although this number is small compared to the number answering in accordance with their first-hand observation, it shows that children do sometimes base their answer on the reliable informant, even in such a simple task. Nevertheless, children's general resistance to inaccurate information should be underlined in one further respect. The search task was designed to clarify the answers of those children who 'followed' the reliable puppet. It is interesting to note that, of the eight children who agreed with the reliable puppet, only two did not look into the box when asked to retrieve the pompon which they had seen the experimenter put into the box. The six other children, even though they had just claimed that the colour of the pompon in the box was of a different colour, still looked into the box when asked to retrieve the pompon of the 'actual' colour. By implication, these children were not completely misled by the reliable puppet even if they had repeated what he had said.

In conclusion, our studies were designed to analyze the ontogenesis of children's trust in testimony. Several interesting results emerged. First, three-year-olds had difficulty in grasping the connection between a potential informant and a piece of information that this informant possessed. To explain this difficulty, we proposed two possible explanations, one focusing on young children's difficulty with deviant modes of communication and the other focusing on their well-known difficulties in conceptualizing false beliefs. Second, a clear limit to children's credulity has been highlighted. Even when children are able to differentiate between reliable and unreliable informants, they do not blindly follow what is said by an apparently reliable informant. On the contrary, most children who are able to distinguish the two sources maintain a certain level of scepticism: they agree with the trustworthy informant only if the communicated information does not contradict with what they already know. Their readiness to reject a false statement leads us to postulate the presence of a filtering mechanism. This mechanism might be related to a specific cognitive module selected to process language inputs (Sperber and Wilson, 2002). This does not mean that cooperation is not important for children during the process of knowledge acquisition—on the contrary. Nonetheless, the confidence that children have in testimony is moderate and we can reasonably describe their stance as one of sceptical trust.

Finally, we may ask whether children's scepticism is ever suspended. After all, if children are not 'empty containers' passively filled up by cultural contents, they are nevertheless confronted with certain kinds of information that are held for true by their community (religious beliefs, scientific facts) without being fully consonant with the knowledge that children already possess. In this context, our study opens up some interesting questions. Recall that, even for sources devoid of authority

(frog and mouse puppets), a certain number of children responded according to what the reliable source had said, even if it contradicted what they had just seen. Would more children display what we have called gullible trust if the informant were not just reliable but also endowed with authority? One could imagine an experimental setting where incorrect information is offered not by a stuffed animal, but by a teacher, for example. In addition, complementary experiments centred on the *content* of the information transmitted may be envisaged. In our studies, the conflict between different channels involved simple perceptual information, namely colour. We can ask what would happen when the contradictory information concerns more complex information, such as the causal knowledge that young children possess in the domain of naïve physics. When combined with the literature on the way that children acquire knowledge 'by themselves', these studies may help in the construction of a more general theory of belief acquisition in which trust in testimony plays a key role.

Human Development and Psychology
Graduate School of Education
Harvard University

References

Astington, J.W., Harris, P.L. and Olson, D.R. (eds.) 1988: *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Baillargeon, R. 1993: The object concept revisited: New directions in the investigation of infants' physical knowledge. In C.E. Granrud (ed.), Carnegie Symposium on *Cognition: Visual Perception and Cognition in Infancy*. Hillsdale, NJ: Erlbaum.

Carey, S. 1995: On the origin of causal understanding. In D. Sperber, D. Premack and A.J. Premack (eds.), *Causal Cognition. A Multidisciplinary Debate*. Oxford: Clarendon Press.

Carey, S. 2001: Evolutionary and ontogenetic foundations of arithmetic. *Mind & Language*, 16, 37–55.

Cosmides, L., and Tooby, J. 1989: Evolutionary psychology and the generation of culture. Case study: A computational theory of social exchange. *Ethology and Sociobiology*, 10, 51–97.

Gelman, R., and Gallistel, C.R. 1978: *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gilbert, D.T. 1991: How mental systems believe. *American Psychologist*, 46, 107–119.

Gilbert, D.T. 1993: The assent of man: mental representation and the control of belief. In D.M. Wegner and J.W. Pennebaker (eds.), *Handbook of Mental Control*. Englewood Cliff: Prentice-Hall.

Giménez, M. and Harris, P.L. 2001: Intimations of immortality and omniscience in early childhood. Paper presented at the Biennial meeting of the Society for Research in Child Development, 19–22 April, Minneapolis.

Harris, P.L. 2002: What do children learn from testimony? In P. Carruthers, M. Siegal and S. Stich (eds.), *Cognitive Bases of Science*. Cambridge: Cambridge University Press.

Hirschfeld, L.A. 1996: *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds*,. Cambridge MA: MIT Press.

Kaufmann, L., and Clément, F. 2003: La sociologie est-elle un savoir infus? Pyramide des sciences et architecture cognitive. *Intellectica*. 36/37, 421–457.

Koenig, M.A., and Echols, C.H. 2003: Infants' understanding of false labeling events: the referential roles of words and the speakers who use them. *Cognition*, 87, 179–208.

Krebs, J.R., and Dawkins, R. 1984: Animal signals: mind-reading and manipulation. In J.R. Krebs and N.B. Davies (eds.), *Behavioral Ecology*. Sunderland, Mass.: Sinauer Associates.

Medin, D., and Atran, S. 1998: *Folk Biology*. Cambridge, MA: MIT Press.

Millikan, R. 1998: A common structure for concepts of individuals, stuffs, and real kinds: more Mamma, more milk, and more mouse. *Behavioral and Brain Sciences*, 21, 55–100.

Mitchell, P., Robinson, E.J., Nye, R.M. and Isaacs, J.E. 1997. When speech conflicts with seeing: young children's understanding of informational priority. *Journal of Experimental Child Psycholog*, 64, 276–294.

Pea, R.D. 1982: Origins of verbal logic: spontaneous denials by two- and three-year-olds. *Journal of Child Language*, 9, 597–626.

Perner, J. 1988: Developing semantics for theories of mind: from propositional attitudes to mental representation. In J.W. Astington, P.L. Harris and D.R. Olson (eds.), *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Reid, T. [1785]1983: *Philosophical Works*. Hildesheim: Georg Olms Verlag.

Robinson, E.J., Champion, H., and Mitchell, P. 1999: Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, 35, 535–546.

Robinson, E.J., Mitchell, P. and Nye, R.M. 1995: Young children's treating of utterances as unreliable sources of knowledge. *Journal of Child Language*, 22, 663–685.

Robinson, E.J. and Whitcombe, E.L. 2003: Children's suggestibility in relation to their understanding of sources of knowledge. *Child Development*, 74, 48–62.

Spelke, E.S. 1991: Physical knowledge in infancy: Reflections on Piaget's theory. In S. Carey and R. Gelman (eds.), *The Epigenesis of Mind: Essays on Biology and Cognition*. Hillsdale, NJ: Erlbaum.

Sperber, D. 2000: Metarepresentations in an evolutionary perspective. In D. Sperber (ed.), *Metarepresentations*. Oxford: Oxford University Press.

Sperber, D. 2001: An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29, 401–413.

Sperber, D., and Wilson, D. 2002: Pragmatics, modularity and mind-reading. *Mind & Language*, 17, 3–23.

Wellman, H.M. 1990: *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Wellman, H.M., Cross, D. and Watson, J. 2001: Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72, 655–684.

Whiten, A., and Byrne, R.W. 1991: The emergence of metarepresentation in human ontogeny and primate phylogeny. In A. Whiten (ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*. Cambridge, MA: Basil Blackwell, Inc.

Wynn, K. 1992: Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220–251.