

**The emergence of selective trust in early childhood**

Paul L. Harris, Elisabeth S. Pasquini, Kathleen Corriveau,

Harvard University

Melissa Koenig

University of Chicago

&

Fabrice Clément

University of Lausanne

To appear in:

*From Metacognition to Self-Awareness*

Edited by Joelle Proust, Jerome Dokic and Elisabeth Pacherie

Oxford University Press

Send correspondence to: Paul L. Harris, Graduate School of Education, Harvard University, 503A Larsen Hall, Appian Way, Cambridge, MA, 02138, USA. (e-mail: Paul\_Harris@gse.harvard.edu).

### Abstract

Children cannot check the veracity of many of the claims they hear. When told the name of an unfamiliar object, or presented with various historical, scientific or metaphysical claims, they typically have to accept such information on trust. Nevertheless, on those occasions when they can check an informant's claims against known facts or prior assertions, children are in a position to form an assessment of the trustworthiness of their informant. In a series of experiments, we show that preschool children aged 3- and 4 years monitor informants' reliability in this fashion. When a particular informant is accurate children prefer to seek and accept information from her rather than from someone who has proved inaccurate. Among 4-year-olds, this monitoring for reliability moderates children's trust in familiar as well as unfamiliar informants.

## **The emergence of selective trust in early childhood**

Human culture is crucially dependent on the transmission of information from one member of the community to another, both within and across generations. A key psychological issue in that transmission process is the degree to which recipients trust informants to supply accurate information. In this paper, we discuss the ontogenesis of selective trust in one informant as compared to another. Our main conclusion is that even preschool children are surprisingly sensitive to variation among informants in their reliability. For the present, we do not understand the exact mechanism that they deploy to make such assessments, but we are beginning to delimit its effects and the conditions under which it operates. We start by describing two experiments showing that preschoolers trust knowledgeable as compared to apparently ignorant informants. We then consider whether young children trust an informant who has been accurate over one who has been inaccurate. Next, we discuss how well children can differentiate between informants when each has been inaccurate, but to varying degrees. Finally, we ask whether children factor in the familiarity of an informant and how they react when a familiar informant proves to be unreliable. In a concluding section, we indicate several directions for further research: experiments that would help to uncover exactly how children gauge an informant's reliability and experiments that would link the ontogenesis of trust to wider questions about children's early attachments and relationships as well as the process of impression formation among adults.

### Ignorance versus knowledge

To explore young children's selective trust, we have used a relatively simple paradigm in which children are first given information about the differential knowledge

of two informants in a familiarization period. Then, in a test period, children are introduced to an unfamiliar object or situation and given an opportunity to seek and accept information about it from one of the two informants. We measure the extent to which children choose to rely on the more knowledgeable of the two informants.

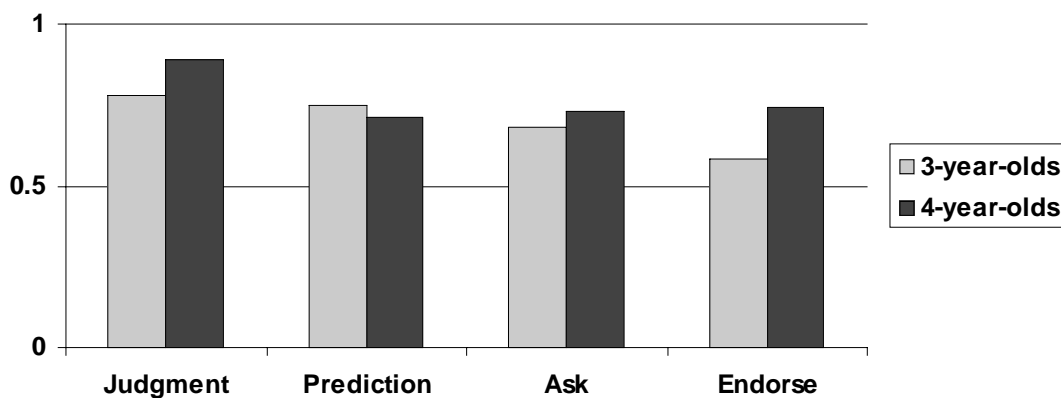
Recent research on children's understanding of mental states shows that even 3-year-olds have some understanding of knowledge as compared to ignorance. For example, 3-year-olds realize that someone who has looked inside a box will know its contents whereas someone who has not looked inside will remain ignorant (Pratt & Bryant, 1990). However, such variation between individuals simply reflects their differential perceptual access to a specific scene or entity (Miller, 2000). We asked whether children are also able to infer more stable and wide-ranging variation between individuals (Koenig & Harris, 2005; Experiment 2). In the familiarization period, children were introduced to two informants – one named each of three common objects accurately whereas the other admitted to not knowing their names. Because the objects were familiar, the children themselves knew their names. In principle, therefore, they could easily recognize that one informant did know the right names for the objects even if the other claimed ignorance.

Before and after the ensuing test period, children were asked to make an explicit judgment about the relative knowledge of the two informants. More specifically, they were asked to judge who was “not very good at answering the questions” about the names of the objects. In the intervening test period, children were first shown another familiar object and asked to predict what the two informants would say about it. They were then shown three novel objects whose names they did not know, invited to ask one

of the informants what each novel object was called, and after each had suggested a different name for the novel object, to say which of the two supplied names was the one that they would use.

Overall, both age groups proved to be remarkably good at monitoring, predicting and utilizing the difference between the two informants (see Figure 1). Thus, in answering the judgment questions, children reliably picked out the informant who was “not very good” at answering the questions. In the prediction trials, they anticipated that one informant would name the object accurately whereas the other would acknowledge ignorance or make a mistake. When given an opportunity to ask for information, they preferred to turn to the knowledgeable as opposed to the ignorant informant. Finally, when given an opportunity to endorse the name supplied by one informant or the other, they tended to endorse the name supplied by the knowledgeable informant – although this selective pattern of endorsement was significant for 4-year-olds but not for 3-year-olds.

Figure 1. Proportion of correct responses as a function of question type and age (Harris & Koenig, 2005, Experiment 2).



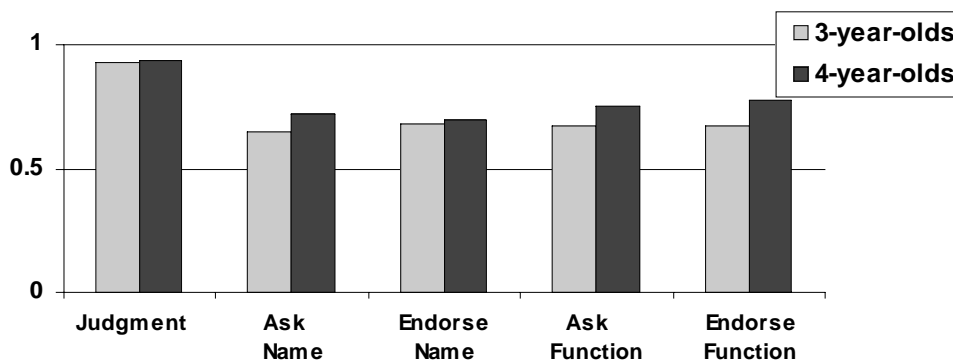
These results imply that children rapidly form an assessment of the knowledge of one informant as compared to another. A considerable body of research suggests that children rarely use trait terms to describe and explain people's actions until middle childhood (Livesley & Bromley, 1973; Yuill, 1992). By contrast, our findings show that in the domain of epistemic reliability, at least, preschoolers are quite quick to detect regularities in the behavior of people they have just met and they use those regularities to anticipate people's future behavior.

In the experiment just described, the two informants varied in the way that they named objects – one was knowledgeable whereas the other admitted ignorance. Subsequently, children tended to trust the knowledgeable informant concerning the names of unfamiliar objects. Was children's selective trust confined to object naming – the domain in which the two informants had varied? Alternatively, would children also display selective trust if the two informants offered information about object functions and not just about object names? This issue was pursued in a follow-up experiment (Koenig & Harris, 2005; Experiment 3).

Children were again introduced to two informants, one who proved accurate and one who proved ignorant in naming familiar objects. Children were then shown four unfamiliar objects. For two of the four objects, they were invited to seek help concerning their names from either of the two informants. For the other two objects, they were invited to seek help concerning the functions of the objects. Figure 2 displays the findings for 3- and 4-year-olds. In the judgment trials, children in both age groups were again very accurate in picking out the person who was “not very good” at answering the questions. As in Experiment 1, they also tended to seek help concerning the names of the unfamiliar

objects from the knowledgeable rather than the ignorant informant – and when both offered a suggestion they typically endorsed the name offered by the knowledgeable informant. Finally, they displayed a very similar pattern with respect to object functions – they preferred to ask for help from the knowledgeable informant. In addition, when the two informants demonstrated a different use for the object children typically endorsed the function modeled by the more knowledgeable informant.

Figure 2. Proportion of correct responses as function of question type and age (Koenig & Harris, 2005, Experiment 3).



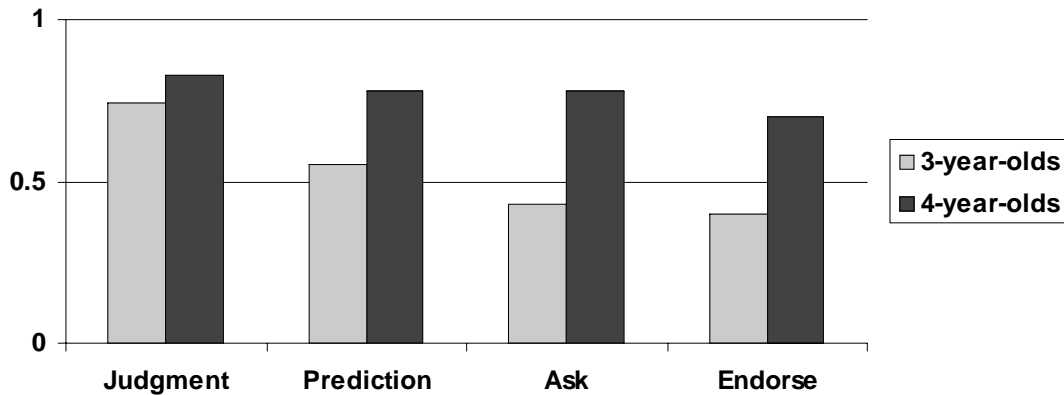
By implication, having learned about the accuracy with which the knowledgeable informant could name objects, children did not make a very specific and narrow assessment of her knowledge. They took her to be knowledgeable about object functions as well as object names. Just how broad children's attributions are remains to be established by future research. Children may have assumed that the knowledgeable informant was better informed in other areas of knowledge as well – such as object locations or animal characteristics. Indeed, they may even have assumed that the

knowledgeable informant was superior not just with respect to knowledge but with respect to other positive traits, such as kindness or motor skill.

#### Accuracy versus inaccuracy

Selective trust might be quite easy for young children to display when they are confronted by an ignorant as compared to a knowledgeable informant – especially when one informant explicitly admits ignorance, as in the two experiments just described. How do they react when the two informants vary in a less explicit fashion? To explore this issue, we have conducted several studies in which children are introduced to one informant who is accurate and another who is inaccurate. In each experiment, the two informants differed in the accuracy with which they stated the name or the properties of three familiar objects (Clément, Koenig & Harris, 2004, Experiments 1 & 2; Koenig, Clément & Harris, 2004; Koenig & Harris, 2005, Experiment 1). In our initial experiments, the contrast between the two informants was quite stark: one was consistently accurate, whereas the other was consistently inaccurate. However, the inaccurate informant did not acknowledge her errors, and so it was up to the child to detect them and to make relevant attributions. In the test phase, the two informants supplied information about the names or properties of unfamiliar objects. In these experiments, the typical pattern was for 4-year-olds to display selective trust by asking for and endorsing information from the accurate informant, but 3-year-olds were often indiscriminate – they tended to ask for and to accept information just as readily from both informants. Figure 3 presents illustrative findings from Koenig and Harris (2005; Experiment 1).

Figure 3. Proportion of responses directed at the more reliable as opposed to the less reliable informant as function of question type and age (Koenig & Harris, 2005, Experiment 1).



Pooling the results of these various experiments, there appears to be an age change. Three- and 4-year-olds generally display selective trust when they encounter two informants – one who is evidently knowledgeable and one who admits to ignorance. In addition, 4-year-olds display selective trust when they encounter an accurate and an inaccurate informant. By contrast, 3-year-olds are less selective. How should this age change be explained? One plausible explanation may be derived from the well-established finding that 3-year-olds are considerably poorer than 4-year-olds at identifying when someone holds a false belief and at recognizing its implications for what that person will say and do (Wellman, Cross & Watson, 2001). Suppose that 3-year-olds were puzzled by the false claims made by the inaccurate informant whereas 4-year-olds were able to attribute them to the informant's false beliefs. In that case, 3-year-olds might not expect her inaccuracy to persist whereas 4-year-olds might plausibly do so.

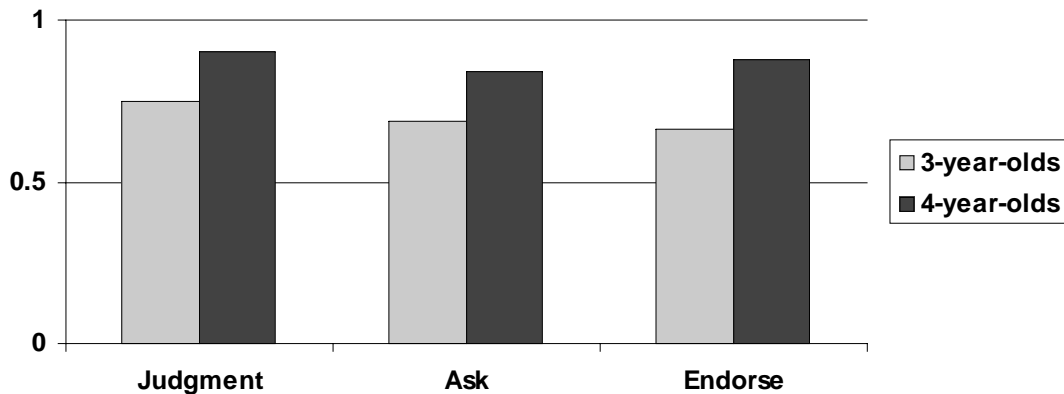
Further inspection of Figure 3 reveals a finding that recurred across this set of experiments and is consistent with the hypothesis that 3-year-olds are puzzled by the informant's inaccuracies. When asked to predict how the inaccurate informant would name or describe a familiar object, three-year-olds, unlike 4-year-olds, often inappropriately claimed that the hitherto inaccurate informant would name or describe it correctly. Notice, however, that 3-year-olds were not completely insensitive to differences between the two informants – when asked for their judgment, they did tend to pick out the inaccurate informant as being “not very good” at answering questions. Recall also that 3-year-olds were selective when faced with a choice between an accurate and an ignorant informant. Overall, then their pattern of responding suggested that they had a conceptual difficulty in interpreting and predicting the informant's inaccuracies but not in recognizing them as such.

At first this explanation appeared to be plausible but subsequent research has led us to question whether 3-year-olds really do have a conceptual difficulty in interpreting and predicting inaccuracy. It seems increasingly likely that 3-year-olds are simply less efficient at monitoring and predicting who is the more accurate informant but not completely incapable. After all, the task is not a simple one. In the first place, children need to differentiate between the informants' claims, noting that one is accurate and the other not. Second, they need to differentiate between informants, encoding the identities of the accurate versus the inaccurate informant. Third, they need to extrapolate from those respective past histories to each person's future accuracy. Three-year-olds might be less efficient than 4-year-olds at any one – or more – of these three steps: accuracy checking, differentiating between informants, and extrapolation.

To increase the likelihood that 3-year-olds might manage all of these three steps more efficiently, we altered the test procedure (Pasquini, Corriveau & Harris, 2006, Experiment 1). First, the number of familiarization trials was increased from three to four – with the accurate informant naming all four objects correctly and the inaccurate informant naming all four incorrectly, thereby giving children an additional opportunity to detect and note differences in the accuracy of the claims. Second, the two informants were made more distinctive from one another in order to facilitate their differentiation and identification. Third, the two informants remained in a fixed position for both the familiarization period and the test period in order to facilitate extrapolation from one phase of the experiment to the next.

These three modifications proved to be quite helpful to 3-year-olds. As Figure 4 shows, 3-year-olds continued to perform less accurately than 4-year-olds but nevertheless, they performed above chance on three types of probe: they explicitly judged the inaccurate informant to be “not very good” at answering questions during the familiarization period and they asked for help from, and endorsed the help provided by, the accurate as opposed to the inaccurate informant.

Figure 4. Proportion of correct responses by age and type of question (Pasquini, Corriveau & Harris, 2006, Experiment 1; 100% vs. 0% condition)



An additional feature of this experiment was that children were assessed for their ability to solve a standard false belief task involving a misleading container. Overall, children performed quite poorly on this task; 3-year-olds performed below chance whereas 4-year-olds performed at chance. However, as just noted, this did not prevent either group from displaying selective trust in the more accurate informant. A clear implication of this conjunction of findings is that correct performance on a standard false belief task is *not* a prerequisite for selective trust in a more accurate informant. Thus, even though 3-year-olds display less selectivity than 4-year-olds, it is unlikely that the age change in that selectivity is due to an improvement in false belief understanding.

Summing up the findings so far, it appears that preschoolers are quite sensitive to variation between informants in their trustworthiness. If one informant is consistently accurate, but the other is either consistently ignorant or inaccurate, 4-year-olds display selective trust. They appropriately judge one informant to be better at answering questions; they anticipate how each informant will describe an unfamiliar object; they

seek information from the more reliable informant; and they selectively endorse the information that they receive from that informant. Three-year-olds display the same pattern when confronted by a knowledgeable as compared to an ignorant informant. They are less selective when differentiating between informants in terms of their relative accuracy. Nevertheless, when given repeated evidence of the accuracy of one informant and the inaccuracy of the other, 3-year-olds are also selective. We are not yet sure why 3-year-olds are less selective than 4-year-olds but for the time being we have no evidence that the age change is linked to their improvement on false belief tasks.

#### Assessing relative accuracy

In the experiments described so far, each of the two informants behaved in a consistent fashion. One was consistently accurate across trials whereas the other was consistently ignorant – or consistently inaccurate – across trials. Outside of the laboratory, however, informants are rarely so consistent. They are likely to display a mix of accuracy and inaccuracy, or truth and error. Despite this mix, we nonetheless judge some informants to be generally reliable whereas we are dubious about others. By implication, we appear to form a global impression of someone's trustworthiness – weighing their overall accuracy against their occasional inaccuracy.

Do preschoolers display a similar tendency? More specifically, when they are faced with informants who are less than fully consistent, do they form a global impression of their trustworthiness? To explore this issue, we included two further conditions in the experiment just described. Recall that in one condition children were introduced to one informant who was accurate across all four trials and one informant who was inaccurate across all four trials. We may refer to this as the '100% vs. 0%'

condition. In two further conditions, we made either the accurate or the inaccurate informant less than fully consistent. Thus, in one condition ('75% vs. 0%') one informant was accurate on three of the four trials whereas the other was consistently inaccurate. In a second condition ('100% vs. 25%'), one informant was consistently accurate whereas the other was inaccurate on three of the four trials.

The results for the 4-year-olds were surprisingly straightforward. In both conditions, they were systematic across all three types of test question. Thus, they picked out the person who was "not very good" at answering the questions, they sought help from the other informant, and when given suggestions by both informants, they typically endorsed what the more accurate informant had said. These results clearly show that 4-year-olds 'forgive' occasional errors. Even though the more accurate informant was not consistently accurate in the 75% vs. 0% condition, 4-year-olds appeared to overlook or forgive that error and treat her as the more trustworthy informant. Moreover, even though the less accurate informant was not consistently inaccurate in the 100% vs. 25% condition, 4-year-olds still kept in mind her errors and treated her as the less trustworthy informant. Stated simply, 4-year-olds appear to recognize and accept that informants will display occasional inconsistency – they will sometimes be accurate and sometimes inaccurate – but that does not prevent them from judging the global reliability of an informant and preferring those who, on balance, are accurate.

The results for the 3-year-olds were less simple but provocative. First, in the 100% vs. 25% condition, although they were somewhat less accurate than 4-year-olds, they were systematic in their answers to all three types of questions. They identified the person who was "not very good" at answering the questions, sought help from the

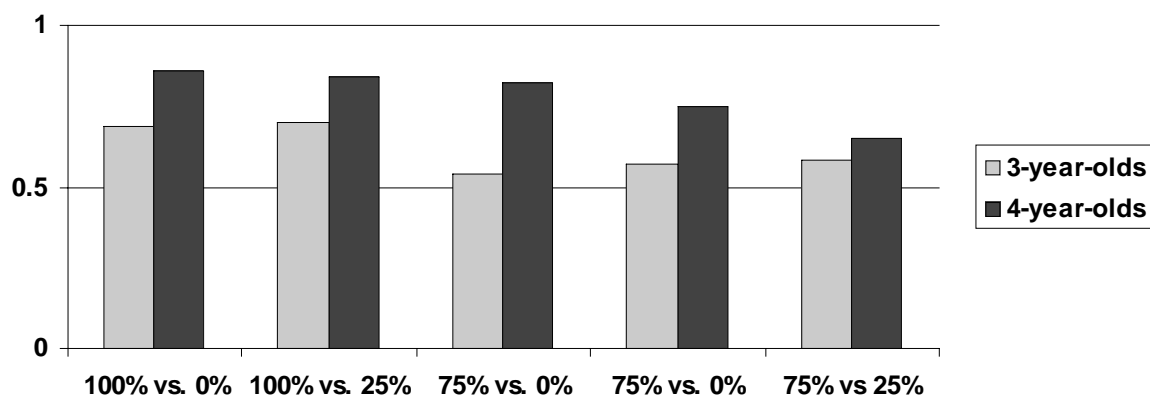
consistently accurate informant and they typically endorsed what the consistently accurate informant said. Thus, 3-year-olds behaved in the 100% vs. 25% condition in much the way as they had behaved in the 100% vs. 0% condition.

On the other hand, in the 75% vs. 0% condition, their response pattern was very different. They behaved in an essentially random fashion across all three test questions. Thus, they failed to identify the informant who was “not very good” at answering the questions, sought help from both informants indiscriminately, and endorsed either informant in a non-selective fashion. Note that the more accurate informant in this condition made a single error – whereas in the other two conditions described above (100% vs. 0% and 100% vs. 25%), the more accurate informant had made no errors. By implication, 3-year-olds are ‘unforgiving’ of even a single error. They treat such an informant as no more trustworthy than someone who has made multiple errors.

Further evidence for the different stance of 3- and 4-year-olds emerged in a follow-up experiment in which children were presented with two conditions: 75% vs. 0% and 75% vs. 25% (Pasquini et al., 2006, Experiment 2). If 4-year-olds can monitor the overall balance of accuracy versus inaccuracy, they should display selective trust in both of these conditions. On the other hand, if 3-year-olds are unforgiving of single errors, they should fail to display selective trust in either condition. The results were in line with these expectations. Overall, 4-year-olds displayed selectivity in both conditions whereas 3-year-olds did so in neither. Figure 4 shows the results (collapsed across judgment, ask and endorse trials) for the three conditions of the initial study (100% vs. 0%; 100% vs. 25%; 75% vs. 0%) and the two conditions of the follow-up study (75% vs. 0%; 75% vs. 25%). Inspection of Figure 5 confirms that 4-year-olds performed above chance in all

conditions, whereas 3-year-olds only performed above chance in only two conditions – those in which one informant was 100% accurate.

Figure 5: Proportion of correct responses by age and condition (Pasquini, Corriveau & Harris, 2006, Experiments 1 and 2)



The two experiments just described produced an intriguing pattern. We can summarize the pattern by saying that 4-year-olds display selective trust when one informant is, on balance, more accurate than another; 3-year-olds are also selective – albeit less so than 4-year-olds – whenever they are able to pick out a consistently accurate informant. If, however, an informant makes a single error, 3-year-olds treat that informant with as much suspicion as someone who has erred consistently. Moreover, as our post-hoc analyses revealed it did not matter to 3-year-olds when the single error occurred in the sequence of trials.

What mechanism might be at work in each age group? We may speculate that children are initially disposed to treat all new informants with zero mistrust. Then, if the

informant subsequently admits ignorance or makes an error, that default level of mistrust is reduced from zero to minus one. Among 4-year-olds, the detection of further ignorance or error leads to a cumulative deepening of mistrust so that after 4 consecutive errors the depth of mistrust would be reduced to minus four. On this argument, 4-year-olds would be expected to display selective trust in all conditions because their level of mistrust in the less accurate informant would be consistently deeper than their level of mistrust in the more accurate informant who would still enjoy either zero mistrust – in the wake of consistent accuracy – or only slight mistrust – in the wake of a single error.

Suppose that 3-year-olds are, in certain respects, similar to 4-year-olds. They too are initially disposed to approach all informants with zero mistrust. Moreover, as soon as an informant acknowledges ignorance or says something that is evidently false, their level of trust decreases from zero to minus one. However, suppose that, unlike 4-year-olds, 3-year-olds do not keep track of further unreliability by an informant in a cumulative fashion: effectively they display a binary stance toward informants of either zero mistrust or (maximal) mistrust of minus one. On this account, 3-year-olds would discriminate between a consistently accurate informant and one who makes one or more false claims. On the other hand, they would treat all informants who make one or more false claims with equivalent mistrust. Thus, as Table 1 spells out, 3-year-olds would show selective trust in two conditions: 100% vs. 0% and 100% vs. 25% but indiscriminate behavior in two other conditions: 75% vs. 0% and 75% vs. 25%. This corresponds, of course, to the pattern of results that was obtained.

Table 1: Hypothetical depth of mistrust, and presence or absence of selective trust, as a function of age and condition

	3-year-olds	Selectivity?	4-year-olds	Selectivity?
100% vs. 0%:	0 vs. -1	Yes	0 vs. -4	Yes
100% vs. 25%:	0 vs. -1	Yes	0 vs. -3	Yes
75% vs. 0%:	-1 vs. -1	No	-1 vs. -4	Yes
75% vs. 25%:	-1 vs. -1	No	-1 vs. -3	Yes

Armed with this tentative model, we turn next to a hitherto unexplored question. How do young children approach a familiar informant? Does familiarity provoke trust?

#### Children's trust in familiar informants

The mechanism that we have tentatively proposed yields the following highly counter-intuitive prediction. Consider the following common situation: the child meets two potential informants, one is a familiar caregiver and the other is a stranger. The stranger will activate zero mistrust because the child has no record of past unreliability for that person. On the other hand, the familiar caregiver will activate some minimal level of mistrust. After all, in the past, he or she will occasionally have produced false claims. Indeed, if we take seriously the behavior of the 3-year-olds in the experiments described so far, one error might be sufficient to dispel any preference for an otherwise reliable caregiver. Recall that 3-year-olds showed no preference for an informant who had been accurate on three out of four trials as compared to an informant who had not been accurate at all.

Thus, if the model that we have described so far is correct, we would expect both 3- and 4-year-old children to be more likely to seek and endorse information from the stranger rather than the familiar caregiver. Indeed, if 4-year-olds keep track of cumulative error – as proposed above – their preference for an unfamiliar informant should be even stronger than that displayed by 3-year-olds.

Given the radical implausibility of each of these two predictions, we need to step back and re-consider the assumptions made so far. On reflection, it is likely that our emphasis on children’s sensitivity to the past inaccuracy – or ignorance – of an informant is too narrow. Presumably, preschoolers keep some kind of record or tally not just of the past *unreliability* of an informant but also of his or her *reliability*. Suppose, as before, that 4-year-olds approach everyone with a default setting of zero. However, on each occasion that the informant makes a true claim, the existing setting is raised by one. If, however, the informant makes a false or ignorant assertion, any accumulated trust is lowered by one. So, for example, following exposure to an informant who makes three true claims and one false claim, the resultant score would be plus two (i.e.,  $0 + 1 + 1 + 1 - 1 = 2$ ). Following exposure to an informant who makes one true claim followed by three false claims, the final score would be minus two (i.e.,  $0 + 1 - 1 - 1 - 1 = -2$ ). Overall, this mechanism would deliver the pattern of scores shown in the right-hand panel of Table 2. Like the model set out in Table 1, it predicts that 4-year-olds will display selective trust across all four conditions.

Table 2: Hypothetical level of trust, and presence or absence of selective trust, as a function of age and condition

	3-year-olds	Selectivity?	4-year-olds	Selectivity?
100% vs. 0%	1 vs. 0	Yes	4 vs. -4	Yes
100% vs. 25%	1 vs. 0	Yes	4 vs. -2	Yes
75% vs. 0%	0 vs. 0	No	2 vs. -4	Yes
75% vs. 25%	0 vs. 0	No	2 vs. -2	Yes

By contrast, suppose that 3-year-olds adopt a simpler, binary classification: if someone has been accurate, the default is raised from zero to (a maximum of) plus one (i.e.,  $0 + 1 = 1$ ). On the other hand, if someone has produced an inaccurate or ignorant claim, any score of plus one is deleted and replaced by the default score of zero (i.e.,  $0 + 1 - 1 = 0$ ). This would yield the pattern of scores shown in the left-hand panel of Table 2. Taken together, these two heuristics predict that 4-year-olds will display selective trust whenever the relative accuracy of one informant exceeds that of another informant. By contrast, 3-year-olds will display selective trust only when one of the two informants has been consistently accurate.

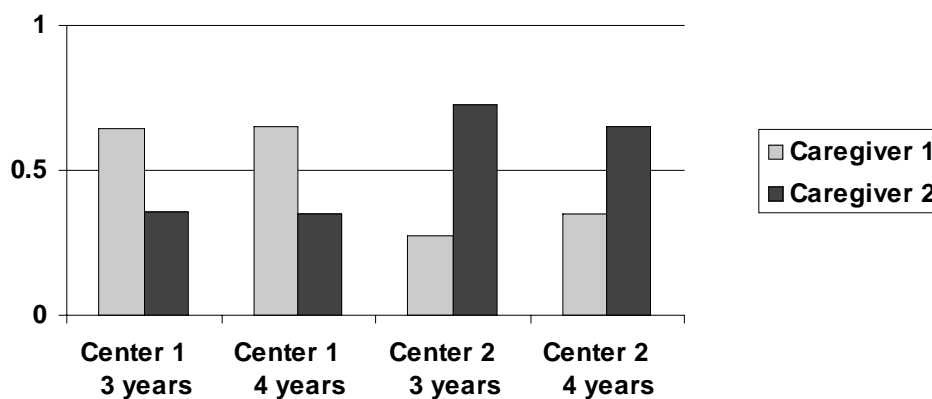
On this model, both 3- and 4-year-olds are also likely to display trust in a familiar informant rather than an unfamiliar informant. Four-year-olds will display such selective trust provided that the familiar informant has been predominantly accurate in the past, whereas three-year-olds will display such selective trust provided that the familiar informant has been consistently accurate in the recent past, i.e., provided some recent error has not led to the deletion and replacement of maximal trust.

Note, however, that neither age group should display unshakeable trust in a familiar informant. More specifically, suppose that children are shown a series of familiar objects and hear a familiar caregiver name them inaccurately and an unfamiliar caregiver name them accurately. Three-year-olds should shift their trust from caregiver to stranger. Even 4-year-olds should eventually shift their trust to the stranger depending on the extent to which they have previously accumulated trust in the caregiver. Conversely, if children hear the familiar caregiver name the objects accurately and the unfamiliar caregiver name them inaccurately, 3-year-olds should retain their initial preference for the more familiar caregiver and 4-year-olds should intensify it.

To examine these predictions, Corriveau and Harris (2006) tested 3- and 4-year-olds for their selective trust in two caregivers, one familiar and one unfamiliar. The Experiment had three phases: pretest trials, accuracy trials and posttest trials. In pretest trials, children were shown 8 unfamiliar objects; they were invited to learn the names of four of these unfamiliar objects and the functions of the remaining four. In both cases, they could seek information from either the familiar or the unfamiliar caregiver. Both caregivers actually supplied information and children were then asked to endorse one or the other. To be sure that children were basing any preference for the familiar caregiver on her familiarity and not on her personal characteristics, children were tested in two affiliated childcare centers. Caregiver 1 worked in Center 1 and was therefore familiar to the children there. Caregiver 2 worked in Center 2 and was familiar to the children there. Figure 6 shows the proportion of trials on which children selected the caregiver from Center 1 or Center 2 as a function of age and Center. Scores are out of a maximum of 16 (i.e., collapsing across Ask and Endorse probes and Object Name and Object Function

trials). Inspection of Table 6 shows that children in each childcare center displayed a clear preference for the more familiar as compared to the less familiar caregiver. This preference was evident among 3- and 4-year-olds alike.

Figure 6. Proportion of times children selected each caregiver on pretest trials by age and childcare center (Corriveau & Harris, 2006).



Children's preference for the more familiar caregiver might, of course, reflect an emotional rather than a cognitive preference. Perhaps children prefer to approach a familiar caregiver because of the emotional bond that they have formed with her or because they are less wary about approaching her than a comparative stranger. This interpretation is certainly plausible for the Ask probe because children had to indicate which informant they would like to ask – even if they did not need to actually formulate and pose a question to her. On the other hand, it is much less plausible for the Endorse probe in which children were asked to choose between the two names or functions that the informants had supplied them with. In addition, as we shall see, 4-year-olds

preference for the more familiar caregiver shifted in the face of reliability information – and this is not what we might expect if the preference were purely emotional.

In the accuracy trials of the experiment, children were shown a set of familiar objects whose names they knew. Half the children heard the familiar caregiver name these objects accurately and the unfamiliar caregiver name them inaccurately. The remaining children heard the reverse arrangement: the familiar caregiver named them inaccurately whereas the unfamiliar caregiver named them accurately. In posttest trials, children were shown four unfamiliar objects and were given Ask and Endorse probes akin to those they had been given in the pretest trials. Thus, we could check whether children's initial preference for the familiar caregiver was either strengthened or undermined depending on her relative reliability in the accuracy trials. The pattern of results differed for the two age groups as illustrated in Figure 7. This shows the proportion of times that 3- and 4-year-olds continued to select the familiar informant both before and after they received information about her relative accuracy during the accuracy trials.

Figure 7. Proportion of times children selected the more familiar informant (collapsing across ask and endorse probes) by age, trial type (pretest versus posttest) and behavior of the familiar informant (accurate versus inaccurate) during the accuracy trials (Corriveau & Harris, 2006).

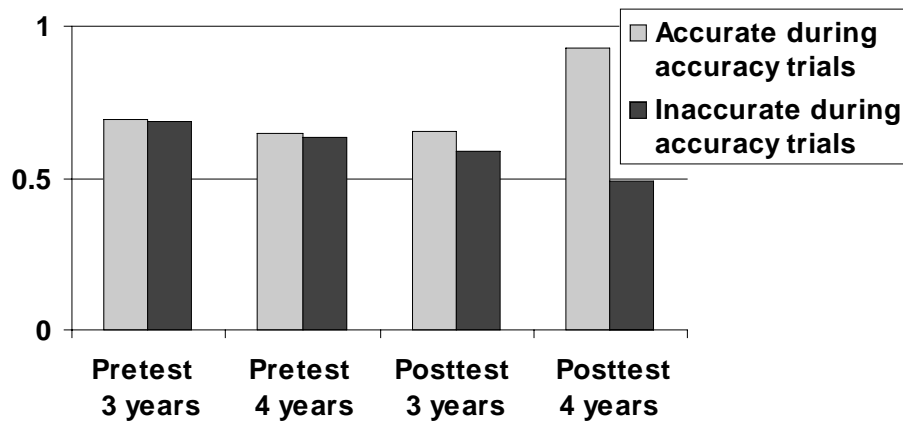


Figure 7 shows that in pretest trials – before children received any information about the relative accuracy of the two caregivers – both 3- and 4-year-olds preferred the familiar caregiver and, not surprisingly, they did so whether or not she would eventually prove to be accurate or inaccurate in the next phase of the experiment – the accuracy trials. This reflects, of course, the selective pattern discussed above. In posttest trials, however, after they had received relative accuracy information, an age difference was apparent. Those 4-year-olds who witnessed the familiar caregiver being more accurate than the unfamiliar caregiver intensified their initial preference for the familiar caregiver whereas those 4-year-olds who witnessed the unfamiliar caregiver being more accurate than the familiar caregiver abandoned that initial preference. Thus, 4-year-olds appeared to fine-tune their preference for a familiar versus an unfamiliar informant depending on their recent history of relative reliability as displayed in the accuracy trials. By contrast,

3-year-olds were less affected by the relative accuracy information. They showed a slight preference for the familiar caregiver if she had been the more accurate but this preference was no stronger than their initial preference.

How do these findings fit the pattern predicted by the model described above? Recall that 4-year-olds should display selective trust whenever the relative accuracy of one informant exceeds that of another informant. The pattern displayed by the 4-year-olds is reasonably consonant with expectation. Suppose, for the sake of argument, that in pretest trials, 4-year-olds start off with a trust level of 8 in the familiar informant and 0 in the unfamiliar informant, and therefore initially display more trust in the familiar informant. Following exposure to the differential accuracy of the two informants, those levels will be altered. If the familiar informant is accurate and the unfamiliar is inaccurate, the respective levels will be 12 and  $-4$ , yielding a still stronger preference for the familiar informant. On the other hand, if the familiar informant is inaccurate and the unfamiliar informant is accurate, the respective levels will be 4 and 4, yielding indiscriminate behavior. This fits the observed pattern although it has to be acknowledged that the alleged fit depends heavily on the default level of trust that is assumed in pretest trials for the familiar caregiver, particularly when the familiar caregiver eventually proves to be inaccurate. If it were set higher, the preference for the familiar caregiver would persist. If it were set lower, it would be trumped by the emerging preference for the unfamiliar caregiver.

Recall that 3-year-olds should display selective trust when one of the two informants has been consistently accurate in the recent past whereas the other has not. Suppose that in pretest trials, 3-year-olds assign the familiar caregiver a trust level of 1

(given her likely pattern of consistent accuracy in the recent past) and the unfamiliar caregiver a trust level of 0. When the familiar caregiver is accurate and the unfamiliar caregiver is inaccurate, those levels will remain unaltered. When the familiar caregiver is inaccurate and the unfamiliar caregiver is accurate, these levels should reverse.

Inspection of Figure 7 provides some support for the first prediction but none for the second.

Summing up, the model offers a satisfactory account of the behavior of 3- and 4-year-olds when an informant's familiarity and her recent accuracy are not set in opposition to one another. It offers a less satisfactory account of cases in which a hitherto reliable informant starts to make errors. In particular, it is as yet unclear how children assign relative weights to past accuracy and recent inaccuracy. We can, however, draw the following important – and plausible – conclusion and then turn to a third model.

The first series of experiments established unequivocally that, when faced with two unfamiliar informants, 4-year-olds and, under certain conditions 3-year-olds, mistrust the less reliable informant – they seek and accept information from the more reliable informant. By contrast, the study just described shows that when 3- and 4-year-olds are faced with a familiar and an unfamiliar caregiver, and the familiar caregiver proves less reliable, neither age group ends up mistrusting her and favoring the less familiar caregiver. Thus, inspection of Figure 7 shows that in the posttest trials, children who witnessed a familiar caregiver making mistakes during the accuracy trials show no strong preference for her but neither do they show any clear-cut signs of mistrust. They approach her with about the same frequency as the unfamiliar (and accurate) caregiver.

Thus, prior familiarity appears to protect an informant's reputation against the otherwise baneful effects of proven unreliability.

Two different interpretations of this protective effect seem feasible. One possibility is that error by a familiar caregiver does not undermine prior trust because the error is re-interpreted. It is seen as an act of pretense or humor – rather than a genuine error. Potential support for this second interpretation emerged when children were asked to say why the informant had made mistakes. Some children said that she was “just pretending.” However, this claim was not significantly more frequent for the familiar as compared to the unfamiliar informant, undermining the hypothesis that children appraise the errors of a familiar informant differently from those of an unfamiliar informant.

An alternative possibility is that an accumulated history of past reliability over weeks or months may produce a relatively deep reservoir of trust. Subsequent short-term inaccuracy may drain that reservoir but not completely. Granted that a familiar caregiver can establish her reliability during many encounters spread over a long period, this is a plausible interpretation. However, it implies that trust in an informant is a graded and cumulative acquisition. Although that implication is quite consistent with the assumptions made about 4-year-olds, it is not easy to reconcile with the assumptions made so far about 3-year-olds. As we have described, some of our findings suggest that 3-year-olds adopt an all-or-none stance rather than a graded stance.

#### First and later impressions

There is a long tradition of research with adults showing that they weigh first impressions more heavily than later impressions. It would not be surprising if preschoolers did the same thing. Consider the following model. In the case of 3-year-

olds, consistent evidence of accuracy during an initial encounter increases trust from 0 to 1 whereas any sign of inaccuracy during an initial encounter blocks that increase.

Conversely, once the default level has been increased from 0 to 1, subsequent encounters that offer evidence of inaccuracy lead to no change in the established level of trust. On this model, 3-year-olds who have come to trust a familiar informant retain that trust despite subsequent inaccuracy on her part.

Four-year-olds might display a more graded version of the same bias toward first impressions. Suppose – retaining but simplifying the assumptions made earlier – that 4-year-olds allocate 5 different levels of trust on first encounter: +2 for consistent accuracy, +1 for predominant accuracy, 0 at baseline, –1 for predominant inaccuracy and –2 for consistent inaccuracy. Suppose, in addition, that once these impressions have been established, they are hard to undermine – they block the uptake of subsequent information about the accuracy or inaccuracy of that person. On this model, children who allocate a level of +2 trust to a familiar informant would retain that level of trust in the face of any accuracy or inaccuracy. This model would yield the pattern of scores shown in Table 3.

Table 3: Hypothetical level of trust, and presence or absence of selective trust, as a function of age, unfamiliarity (U) versus familiarity (F) of informant, and percentage accuracy of informant.

Condition	3-year-olds	Selectivity?	4-year-olds	Selectivity?
U:100% vs. U: 0%	1 vs. 0	Yes	2 vs. -2	Yes
U:100% vs. U:25%	1 vs. 0	Yes	2 vs. -1	Yes
U:75% vs. U:0%	0 vs. 0	No	1 vs. -2	Yes
U:75% vs. U:25%	0 vs. 0	No	1 vs. -1	Yes
F: pre vs. U: pre	1 vs. 0	Yes	2 vs. 0	Yes
F:100% vs. U:0%	1 vs. 0	Yes	2 vs. -2	Yes
F:0% vs. U:100%	1 vs. 1	No	2 vs 2	No

Inspection of Table 3 shows that among 3-year-olds trust would be set at 1 for an unfamiliar informant making no errors but at 0 for an unfamiliar informant making errors. For familiar informants trust would be set at 1 and remain there despite subsequent errors. Among 4-year-olds, trust toward unfamiliar informants would range from 2 to -2, depending on the balance of accuracy to inaccuracy. For familiar informants, trust would be set at 2 and remain there despite subsequent errors. The predicted pattern of selectivity corresponds to the findings reported earlier: the four conditions studied across two experiments by Pasquini et al. (2006) as well as the pre- and posttest trials studied by Corriveau et al. (2006). In sum, a workable model can be constructed if the assumption is made that initial impressions of trustworthiness are resistant to being undermined. The model also implies that unfamiliar informants are rapidly appraised for their reliability –

in a relatively harsh fashion by 3-year-olds and in a more nuanced fashion by 4-year-olds. This model of impression formation is consonant with findings for adults. Thus, a long tradition of research on impression formation has shown that adults form impressions rapidly and lend a disproportionate weight to initial impressions.

### Future Directions for Research

Three broad directions for future research will be discussed. First, it is important to ask about ways in which the experimental paradigm might be expanded to probe real-life situations in which children are likely to encounter false or misleading claims. Second, a key direction for future research is the design of simple experiments aimed at supplying critical information for model building. Finally, recent research with adults has begun to hint at the neural processes involved in trust. We briefly review these findings below and ask about their connections to the development of trust in early childhood.

Recent studies of maternal discourse have underlined the fact that mothers vary considerably in the psychological language that they use. Some mothers are quite sensitive to their child's mental states and talk appropriately and accurately about them. By contrast, other mothers sometimes talk inappropriately about their child's mental states – they misinterpret them – attributing a preference, intention or emotion that close observation suggests the child does not actually have. Such inappropriate comments are relatively uncommon. Nevertheless, they appear to be a good index of mothers' psychological insight. First, infants who have mothers that produce such inappropriate comments are more likely to be insecurely attached (Meins et al., 2002). In addition, mothers who direct more inappropriate comments to their infants at 6 months are less inclined to describe those children in psychological terms at 4 years of age, focusing

instead on their behavioral and physical characteristics (Meins et al., 2003). These findings highlight the fact that even if parental discourse rarely displays the kind of gross inaccuracies that we have used in our two informant tasks, parents are sometimes inaccurate. Thus, they do misdescribe children's psychological states. In future research, it will be interesting to examine children's reactions to an informant who misdescribes a psychological state rather than a physical object as in the studies reported hitherto. For example, it is feasible to imagine an informant who misdescribes the emotional state of a story character or indeed the emotional state of the child being tested. Such inaccuracies might appear to be subtle compared to the blatant misnaming of an object but from a child's point of view they may be especially potent indices of an adult's unreliability.

A further issue arising from the intriguing findings of Meins and her colleagues concerns the broader question of how a parent's discourse influences the nature of the parent-child relationship. As noted, parental inaccuracy in the psychological domain – and, arguably, in other domains as well – is associated with an insecure attachment. Insofar as a child's attachment status is, at least in part, an index of parental reliability in communication, we may expect securely attached children to be more likely to seek and endorse information from the parent. Faced with two informants, one the parent and one a stranger, securely-attached children should display a clear preference for the parent, whereas insecure children might be indiscriminate or even display a preference for the stranger.

Turning to the issue of future model-building, it is worth acknowledging an important ambiguity running through the experiments that have been reported. When children respond selectively to two informants who have previously varied in accuracy

(whether consistently or inconsistently), we cannot independently gauge the impact of informant accuracy as compared to informant inaccuracy. Children might be impressed exclusively by an informant's inaccuracy – taking his or her accuracy for granted – but given the design of our studies so far we cannot tell.

Recent findings strongly suggest that children take both accuracy and inaccuracy into account in line with the assumptions of models 2 and 3, described above. In a clever design, Jaswal and Neely (in press) introduced 3- and 4-year-olds to two informants, one an adult and the other a child. Children watched as these two informants named four objects. In the 'both reliable' condition, each of the informants gave acceptable names for the objects. For example, one called a shoe a *sneaker* and one called it a *shoe*. Following these accuracy trials, children received post-test trials in which the two informants supplied different names for novel objects. In the 'both reliable' condition, children favored the adult. However, in two further conditions only one of the two informants was reliable during the accuracy trials. Children subsequently favored the reliable informant, whether it had been the child or the adult. These results clearly show that children were attentive to inaccuracy. More specifically, the preference for the adult that they had displayed in the *both reliable* condition was eliminated if the adult, but not the child, had been inaccurate. A final condition – in which both informants were unreliable in the inaccuracy trials – also proved informative. Children trusted neither informant systematically, and often said that neither of the novel names they supplied was correct. To the extent that children did trust either the adult or the child if one of them had proved reliable during the accuracy trials, the experiment confirms that children also attended to accuracy.

In future studies, it will be helpful to learn whether the two types of information, accuracy and inaccuracy have an equally potent effect. This question can be tackled in the following way. Suppose that children are introduced to two unfamiliar informants. Half the children then watch while one informant accurately names objects whereas the other simply calls attention to them (i.e., “Oh look at that!”). The remaining children watch while one informant inaccurately names objects whereas the other simply calls attention to them. Then, on test trials, both groups of children have an opportunity to put questions to and endorse information from the two informants. Depending on how far children’s trust is strengthened by accuracy and undermined by inaccuracy, or both, we may expect to observe varying degrees of selective trust in each group.

A second source of uncertainty running through the various models described above concerns the issue of memory. All of our experiments have been conducted over a relatively short period. The child meets two informants, observes their differential reliability for a few minutes, and immediately thereafter is invited to trust one or the other informant. For our findings to be genuinely relevant to children’s information-seeking, it will be important to demonstrate that the effects we observe have some durability over days and even weeks. Moreover, for model-building purposes it will be helpful to find out if the decay rate is equivalent for both accuracy and inaccuracy information. Conceivably, inaccuracy information has a slower decay rate than accuracy information. In that case, of course, the models described earlier would only hold for immediate testing. Delayed testing would increasingly reflect the impact of prior inaccuracy.

Finally, we may consider research on impression formation among adults. It has long been known that adults rapidly attribute traits to other people on the basis of minimal

behavioral information, often ignoring plausible situational explanations (Gilbert & Malone, 1995; Macrae & Bodenhausen, 2000). Recent findings indicate that such trait information is retrieved in a relatively automatic fashion. Thus, when adults are invited to perform a simple face recognition task, those faces that have been previously associated with particular behaviors elicit greater activity in areas of the brain generally associated with social cognition – the anterior paracingulate cortex and the superior temporal sulcus – than do novel faces (Todorov, Gobbini, Evans & Haxby, in press). Moreover, trait information can be recovered even in the absence of memory for the particular behaviors that gave rise to the trait inference (Todorov & Uleman, 2002).

It is plausible that young children are similarly prone to rapid and automatic trait attribution. The experiments reviewed above already indicate that trait attribution is relatively fast. Recall that children watched the two informants name three or four objects over a period of 3-4 minutes. On the strength of that minimal input, they subsequently inferred the informants' reliability not just with respect to object naming but also with respect to object functions. What we do not yet know are the cues to which children 'bind' that trait information. In order to help them distinguish between the informants – who were both women in their 20s – children's attention was repeatedly drawn to the women's differently colored shirts (e.g., "I bet one of these people can help us find out. Which person would you like to ask, C. with the pink shirt, or S. with the black shirt?"). Arguably, however, these experimenter-supplied prompts to attend to distinctive items of clothing were unnecessary, and possibly even disruptive, for children. The research with adults implies that the face is a crucial site for the binding of trait information. This makes sense from an evolutionary point of view because the face is a highly distinctive

and stable marker of a particular individual whereas a given item of clothing is neither highly distinctive (others may wear the same item) nor stable (the person may dress differently from day to day). Thus, the face offers a reliable set of cues for re-identifying a given individual and valid grounds for the re-activation of previously stored trait information when re-encountered. To examine this capacity in the context of our experimental paradigm is relatively easy. During test trials, children can be shown films of the two informants that reveal only their faces rather than their upper body and clothing. In the longer term, it should also be possible to explore how far children, like adults, more or less automatically recover trait information when confronted by a familiar face.

### Conclusions

Research on children's theory of mind has often probed their understanding of psychological regularities that are person-independent – for example, their understanding of the relationship between perceptual access and knowledge or between belief and emotion. When children interpret the behavior of a given agent in light of such stable mental relationships, there is no implication that they are reaching any conclusions about the distinctive mental traits of the individual agent. Yet social interaction frequently does turn on our attribution of distinctive mental traits to individuals. The results we have described offer a glimpse of that attribution process in early childhood. We are able to investigate not just children's competence at making such attributions, we can also track the conditions under which they are led to make those attributions. Equally exciting is the fact that children's attributions of reliability clearly guide their larger epistemic endeavors. Although there is much about the world that children can discover for

themselves, there is also a great deal that they must take on trust. Knowing whom to trust is, therefore, of major importance for cognitive development, even if its role has generally been ignored.

## References

- Clément, F., Koenig, M., & Harris, P.L. (2004). The ontogenesis of trust in testimony. *Mind and Language, 19*, 360-379.
- Corriveau, K., Pasquini, E.S., & Harris, P.L. (2006). Choosing your informant: Weighing familiarity and past accuracy. Weighing familiarity and past accuracy. Paper submitted for publication.
- Gilbert, D.T. & Malone, P.S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38.
- Jaswal, V.K. & Neely, L.A. (in press). Adults don't always know best: preschoolers use past reliability over age when learning new words. *Psychological Science*.
- Koenig, M., Clément, F. & Harris, P.L. (2004). Trust in Testimony: Children's use of true and false statements. *Psychological Science, 10*, 694-698.
- Koenig, M. & Harris, P.L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261-1277.
- Livesley, W.J. & Bromley, D.B. (1973). *Person perception in childhood and adolescence*. London: Wiley.
- Macrea, C.N. & Bodenhausen, G.V. (2000). Social cognition. Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.
- Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*, 1715-1726.
- Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Das Gupta, M., Fradley, E.,

- & Tuckey, M. (2003). Pathways to understanding mind: Construct validity and predictive validity of maternal mind-mindedness. *Child Development, 74*, 1194-1211.
- Miller, S. A. (2000). Children's understanding of preexisting differences in knowledge and belief. *Developmental Review, 20*, 227-282.
- Pasquini, E.S., Corriveau, K., & Harris, P.L. (2006). Three and 4-year-olds keep track of an informant's accuracy. Paper submitted for publication.
- Pratt, C., & Bryant, P. E. (1990). Young children understand that looking leads to knowing (so long as they are looking through a single barrel). *Child Development, 61*, 973-982.
- Todorov, A., Gobbini, M.I., Evans, K.K., & Haxby, J.V. (in press). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*.
- Todorov, A. & Uleman, J.S. (2002). Spontaneous trait inferences are bound to actor's faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*, 1051-1065.
- Wellman, H.M., Cross, D.D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*, 655-684.
- Yuill, N. (1992). Children's conception of personality traits. *Human Development, 35*, 265-279.

There are many occasions on which parental discourse may not match reality as the child sees it. The parent may make a rash promise and fail to keep it; or confidently predict that one event may be enjoyable when it turns out to be boring – or dangerous when it turns out to be uneventful. The parent may assert that X is too big, too expensive, or just too much – when the child thinks it is nothing of the kind. In such cases, we might plausibly expect