Cognition 112 (2009) 367–380

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The moral, epistemic, and mindreading components of children's vigilance towards deception

Olivier Mascaro *, Dan Sperber

Institut Jean Nicod, UMR 8129, Pavillon Jardin, Ecole Normale Supérieure, 29, rue d'Ulm, F-75005 Paris, France

ARTICLE INFO

Article history:
Received 14 September 2008
Revised 15 May 2009
Accepted 18 May 2009

Keywords:
Trust
Deception
Theory of mind
Communication
Cooperation
Folk epistemology

ABSTRACT

Vigilance towards deception is investigated in 3- to-5-year-old children: (i) In Study 1, children as young as 3 years of age prefer the testimony of a benevolent rather than of a malevolent communicator. (ii) In Study 2, only at the age of four do children show understanding of the falsity of a lie uttered by a communicator described as a liar. (iii) In Study 3, the ability to recognize a lie when the communicator is described as intending to deceive the child emerges around four and improves throughout the fifth and sixth year of life. On the basis of this evidence, we suggest that preference for the testimony of a benevolent communicator, understanding of the epistemic aspects of deception, and understanding of its intentional aspects are three functionally and developmentally distinct components of epistemic vigilance.

1. Vigilance towards deception: part of human communicative abilities

Communication gives access to a vast amount of information, much wider than what could have ever been acquired through direct individual learning. While providing extraordinary benefits, communication is also a source of vulnerability to misinformation.

Competent communicators must exert what we propose to call "epistemic vigilance", that is, an ability aimed at filtering out misinformation from communicated contents. How does epistemic vigilance develop in childhood? Our goal here is to help answer this question with special reference to vigilance towards deception.

One way for avoiding misinformation is to trust informants neither blindly nor randomly, but in a way that is sensitive to their knowledge and honesty. Generally, humans rely heavily on two dimensions to characterize other people and predict their behavior: benevolence – their perceived good or ill intentions – and competence – their perceived ability to execute those intentions (for reviews, see Fiske, Cuddy, & Glick, 2007; Wojciszke, 2005). These two dimensions can be seen as critical in identifying good cooperators, that is, people who are willing and able to help. In the case of communication, competent informants are those able to provide relevant information, while benevolent informants are those willing to provide it (Sperber, 1994). Incompetence produces accidental misinformation; i.e. mistakes, whereas malevolence produces intentional misinformation, i.e. deception.

Honest communicators and their audience share a preference for true information. Deception, on the other hand, is intentional and is normally advantageous to the deceiver and costly to the audience. From the point of view of speakers, the possibility of deceiving one's audience and manipulating their beliefs can be seen as an integral part of what makes communication advantageous. From the point of view of the audience, the risk of deception—much more than that of honest mistakes—jeopardizes the advantageousness of communication. The resulting evolutionary paradox is well known: communicating populations face the risk of being invaded by deceivers, leading to the disappearance of communication itself (Dawkins & Krebs,

* Corresponding author. Tel.: +33 1 44 32 29 76.
E-mail address: olivier.mascaro@gmail.com (O. Mascaro).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.05.012

1978; Krebs & Dawkins, 1984). Just as one can infer from the ongoing existence of cooperation among humans that there must exist psychological and/or social mechanisms that thwart cheating to an extent sufficient to keep cooperation advantageous, one can infer from the ongoing existence of communication that there must exist mechanisms that thwart deception and keep communication advantageous (see Bergstrom, Moehlmann, & Boyer, 2006; Sperber, 2001; for animal communication, see Searcy & Novicky, 2006). Epistemic vigilance exerted by the addressees of communication is, we suggest, such a mechanism.

2. A three step model for vigilance towards deception

Despite its theoretical relevance, young children's ability to be vigilant towards lying, as opposed to their ability to lie themselves, has hardly ever been studied. There are studies that are indirectly relevant to the issue and on which we have drawn in designing our own experiments (Couillard & Woodward, 1999; Freire, Eskritt, & Lee, 2004; Lee & Cameron, 2000; Shultz & Cloghesy, 1981); and recent researches have targeted 6- to 10-year-olds' sensitivity to honesty in a particular domain: self-reports (Gee & Heyman, 2007; Heyman, Fu, & Lee, 2007; Heyman & Legare, 2005; Mills & Keil, 2005). Still, the development of the basic mechanisms allowing one to resist deception as such remains to be explored.

Liars have three characteristic features. They are malevolent, that is, willing to harm others. They do so by communicating false information. They are moved in doing so by the intention to deceive their audience. A fully-fledged capacity to be vigilant towards lying should have, then, has three aspects: a moral/affective aspect involved in attending to malevolence; an epistemic aspect involved in attending to falsity; and a mindreading aspect involved in attending to the liar's intention to deceive. Some epistemic vigilance can nevertheless be exerted on the basis of just the first or the first two of these three aspects. In a rudimentary form, vigilance might be based on nothing more than a preference for the testimony of a benevolent informant over that of a malevolent one, without any understanding of the distinctive intentional and epistemic features of deception. In a less rudimentary form, it might also involve the ability to process the testimony of a malevolent informant as false, without however understanding the intention to deceive. Fully-fledged vigilance toward lying involves a grasp of its moral, epistemic and intentional features.

2.1. The moral component

Little is known regarding children's sensitivity to benevolence and malevolence in communication. Only a few experiments on epistemic trust have targeted variables that may affect the assessment of benevolence, such as familiarity (Corriveau & Harris, 2009; Harris, Pasquini, Corriveau, Koenig, & Clement, in press) or attachment relationship to the mother (Corriveau et al., 2009). However, there is good evidence that infants and young children possess an early capacity to distinguish benevolence from malevolence in general. Infants may be sensitive to the difference between an intention to help versus hinder (Kuhlmeier, Wynn, & Bloom, 2003; Premack & Premack, 1997) and have been claimed to use this sensitivity to guide their preferences for interaction by the age of 6 months (Hamlin, Wynn, & Bloom, 2007). 28-month-olds use words referring to goodness and badness (Bretherton & Beeghly, 1982). Around 4 years of age, children have been shown to use such type of broad assessment in verbal tasks to predict behavior (Boseovski & Lee, 2006; Cain, Heyman, & Walker, 1997; Liu, Gelman, & Wellman, 2007), to infer people's emotional states (Heyman & Gelman, 1999), preferences (Heyman & Gelman, 2000) and to evaluate the appropriateness of aggressive behavior (Gilles & Heyman, 2005). It is quite conceivable therefore that young children might use benevolence to adjust their level of trust in testimony from an early age.

2.2. The epistemic component

A more refined stage of vigilance towards deception involves not only the ability to mistrust malevolent people, but also the capacity to treat lies as false (even if, when the beliefs of the liar happen to be false, a lie may be true; see Adler, 1997).

Work done on vigilance towards incompetence throws some light on children's understanding of the epistemic status of communicated information. Around 3 to 4 years of age, children display sensitivity to epistemic modalities (Jaswal & Malone, 2007; Matsui, Miura, & McCagg, 2006) and expression of ignorance (Koenig & Harris, 2005; Sabbagh & Baldwin, 2001; Sabbagh, Wdowiak, & Ottaway, 2003). Preschoolers' level of trust is affected by informants' level of accuracy in labeling objects and functions (e.g. Birch, Vauthier, & Bloom, 2008; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig & Harris, 2007; Scofield & Behrend, 2008), in demonstrating games rules (Rakoczy, Warneken, & Tomasello, in press) or in reporting episodic information (Clément, Koenig, & Harris, 2004; Eskritt, Whalen, & Lee, 2008; Jaswal & Neely, 2006, "quality" condition). Preschoolers trust more testimonies coming from communicators who are more relevant (Eskritt et al., 2008, "relation" and "quantity" condition), better informed (Nurmsoo & Robinson, 2009; Robinson, Champion, & Mitchell, 1998; Robinson & Whitcombe, 2003; Welch-Ross, 1999; Whitcombe & Robinson, 2000) or who are presented as more competent (Fusaro & Harris, 2008; Lampinen & Smith, 1995).

These results suggest precocious abilities to adjust trust according to informant's competence. However, the naïve epistemology underpinning these abilities remains to be explored. In these studies, children are weighing information coming from two different sources, either from two different informants or from one informant and from themselves. They may merely be ignoring or discarding the information provided by the incompetent testifier, or, in a more sophisticated manner, they may judge it to be false. Investigating children's ability to understand the falsity of lies should increase our knowledge of the naïve epistemology involved in early epistemic vigilance.

2.3. The mindreading component

Not only do liars communicate false information, but also they intend to mislead their audiences. Understanding that A is trying to deceive B—as opposed to unintentionally misinforming B—involves attributing to A's the intention to cause B to form a false belief. The liar's intention being metarepresentational, the attribution of such an intention is itself a second-order metarepresentation (Peskin, 1992; Sperber, 2000).

Extensive research has been devoted to children's ability to engage in deception. Studies addressing directly the problem of lie production have however led to somewhat contradictory results (e.g. see Chandler, Fritz, & Hala, 1989; Hala, Chandler, & Fritz, 1991; Sodian, 1991). Prima facie lies have been observed before the age of four, in naturalistic studies, (Dunn, 1991; Newton, Reddy, & Bull, 2000, Study 2; Wilson, Smith, & Ross, 2003) and in experimental settings where children falsely deny that they have done something forbidden (Lewis, Stranger, & Sullivan, 1989; Polak & Harris, 1999; Talwar & Lee, 2002), or provide false information to conform with politeness rules (Talwar, Murphy, & Lee, 2007). Whether children actually lie in all these cases remains controversial: they may well rely on a punishment avoidance procedure (Perner, 1991; Polak & Harris, 1999), comply to a social norm, display pretend play behavior, express wishful thinking, rather than engage in genuinely deceptive behavior. Moreover, the ability to deceive does not require the capacity to understand deceptive intents: young children may intend to deceive without metarepresenting this intent. From this set of studies, it is hard to draw any clear conclusion about the onset of deceptive intents understanding. On the other hand, studies tapping directly the understanding of deceptive intents as manifested in distinguishing lies from jokes suggest a relatively late development of this capacity, not before 5 to 6 years of age (Sullivan, Winner, & Hopfield, 1995; Winner & Leekam, 1991).

3. Testing the model

A series of studies was designed to investigate more precisely the developmental tendencies suggested by the existing literature. Study 1 asks at what age children show a preference for benevolent as opposed to malevolent informants. Study 2 asks at what age children are capable of judging that the claim made by a liar is false and to infer true information from this judgment. Study 3 asks at what age children's vigilance towards lying relies on both the intentional and the epistemic features of deception.

4. Study 1: Choosing an informant on moral grounds

As we mentioned in introduction, there is good experimental evidence that very young children do distinguish 'good' and 'bad', or 'nice' and 'mean' characters. What is not obvious is whether these children would draw on this distinction in deciding whom to believe. To find out we designed an experiment where participants had to choose between the testimony of a 'nice' agent, and that of 'mean' one.

4.1. Method

4.1.1. Participants

Twenty-three young 3-year-olds were enrolled in the study (M = 3;3, range 2;11 to 3;10). They were tested individually in their school.

4.1.2. Design and procedure

For the testimony task, the child and the experimenter were seated facing each other across a small table. A closed container was placed at the center of the table, and two animal puppets were placed on the left and the right side of the container. The participant had to find out what was hidden in the box. The hidden object was selected from two objects of similar size and familiarity (e.g. a fork and a spoon). Choice of the hidden object was systematically varied across participants. The puppets were then introduced. Each puppet was named after the animal it represented, i.e., a frog puppet was named "the frog" and a cow puppet was named "the cow". One was said to be "kind". The experimenter made the puppet caress him and said, for instance, "The cow stroked me!" The other puppet was said to be "mean". The puppet was made to hit the experimenter who said, for instance, "The frog hurt me!" A control question was then asked. The experimenter took each puppet in turn and asked, "Is it kind?" If children failed this control question, the characterization procedure was repeated once. Children failing three times on control questions were excluded from analysis.

Both puppets were then given perceptual access to the content of the container. The container was opened, with its lid masking the contents of the box from the child. The puppets looked in turn into the container. The experimenter checked that the child had been attentive during the process by asking a control question for each puppet ("Did it look inside the box?"). The procedure was repeated if children failed on this control question. Failing twice led to exclusion from analysis. Each puppet in turn jumped close to the child and the experimenter said in a distinctive voice, for instance, "Inside the box, there is a fork". The kind puppet always told the child what was actually in the box (e.g. saying there was a fork in the box), whereas the mean puppet provided a misleading testimony (e.g. saying there was a spoon in the box). The child was then asked the test question, for instance, "So, what is inside the box? A fork or a spoon?" If children said that both objects were in the box, the experimenter would insist: "No. Remember, there can be only one thing in the box. So what is it, a fork or a spoon?". Children received no feedback on whether the characters provided accurate information: Once children had given their answer to the first testimony question, the experimenter said: "We are going to play the same game, but with another box!" The game was then repeated with a new box, and two new alternative objects served as potential contents of the box. At the end of the second test, children were thanked for their participation. They were invited to choose which of the two characters should give them a gift: "Who would you like to give you a present?".

4.2. Results

All statistical tests employed in this paper are two-tailed. Four children were excluded from analysis. Two children failed on the control questions (3;1 and 3;3), one child became fussy (3;5) and one did not want to pursue the experiment till the end (3;3). Within the remaining participants ($n = 19$, $M = 3;3$, range 2;11 to 3;10) performance did not differ significantly in the first and second test trials ($p = 0.34$; McNemar's test.). The score of children over the two tests was thus aggregated in a single index ranging from 0 for two incorrect answers to 2 for two correct answers. Children performed above chance on this measure of performance with a mean score of 1.36 ($W_+ = 40$, $W_- = -5$, $p = 0.023$; Wilcoxon signed rank test). Notwithstanding this good level of performance, the individual strategies of children were rather inconsistent: 10 participants did not trust the same character over the two tasks, but changed their preference (i.e. selected the testimony of the kind character once and the testimony of the mean character once). The overall pattern of answers did not differ from chance ($p = 1$; binomial test).

Children were separated into two groups depending on whether they asked for a present from the kind ($n = 12$, $M = 3;3$, range 2;11 to 3;8) or mean character ($n = 7$, $M = 3;3$, range 3;0 to 3;10). Children who asked for a present from the kind character consistently chose the testimony of the kind character (83% of correct answers on the testimony test, $W_+ = 36$, $W_- = 0$, $p = 0.006$; Wilcoxon signed rank test); they performed significantly better than children who asked for a present from the mean character ($U = 12$, $p = 0.01$; Mann–Whitney test), who were at chance when it came to deciding whose testimony to believe (43% of correct answers on the testimony test; $W_+ = 0$, $W_- = -1$, $p = 1$; Wilcoxon signed rank test).

4.3. Discussion of Study 1

Participants used benevolence to adjust their level of epistemic trust. This result confirms earlier findings showing that young children can identify dispositions toward benevolence – or malevolence – from an early age. Study 1 shows moreover that children rely on this dimension in the domain of communication. By the age of three years, children are capable of giving more credence to the words of a kind interlocutor than to those of a mean one. This, together with earlier work showing that children take into account an informant's competence, establishes that young children are not simply gullible and do adjust their level of epistemic trust in a variety of ways.

Since the capacity to prefer interaction with a benevolent character has been demonstrated in infants, it is likely that those participants who selected the gift of the malevolent testifier were not paying attention to its moral disposition (rather than being unable to make relevant attributions). This would explain why they did not either show any preference for one informant over the other. Conversely, the evidence shows that children who paid attention to the moral dispositions of the characters relied on these dispositions both in the non-epistemic task of choosing a gift from one of them and in the epistemic task of deciding which testimony to trust.

What we have shown so far is that children as young as three display a preference for the testimony of a kind character over that of a mean one. How is this preference to be explained? Does the evidence show that children understand that malevolent informers may intentionally provide them with false information? Do children in all the age groups we tested understand deception? In fact, there is a much more parsimonious plausible explanation of the evidence.

Children's good performance in the benevolence game could be explained by assuming that they are guided by a preference for interaction with benevolent rather than malevolent individuals. Children may have relied on such moral/affective preferences when selecting the testimony of the benevolent character, just as they did when choosing from which character to accept a gift. If this explanation is correct, then what might look like epistemic vigilance is just a by-product of a more general moral preference. This line of interpretation is consistent with recent findings confirming Premack and Premack's conjecture (1997) and showing that infants display approach/avoidance behavior following the observation of actions that can be interpreted as helping or as hindering an agent (Hamlin et al., 2007).

A somewhat less parsimonious variant of this explanation gives a greater role to cognitive assessment of character as opposed to mere affect-based moral preference. Studies on trait attribution have shown that young children attribute all sorts of positive qualities to 'good' individuals (e.g. being smarter) and negative qualities to the 'bad' individuals (e.g. being less intelligent) (Alvarez, Roger, & Bolger, 2001; Cain et al., 1997; Heyman, Gee, & Giles, 2003). It is possible that children extend such positive and negative assessments to an assessment of the value of nice and mean agents as communicators and prefer, on such grounds, to listen to the nice ones. This would not imply that children acting on such preferences make an epistemic evaluation of the two testimonies, or understand that that of the nice agent is more likely to be true. In the same vein, it might also be that, through a kind of halo effect, children consider mean characters as less intelligent and less likely to provide good information, and nice characters as more intelligent and more likely to provide good information. Their preference for the nice character's testimony might then involve a modicum of epistemic assessment, but would still fall short of understanding truth-and-falsity and even more so of understanding deception. Study 2 investigated these hypotheses.

5. Study 2: Treating lies as false

Whereas in Study 1, participants were presented with the testimony of two communicators about the content of one box, in Studies 2 and 3, they were presented with the testimony of a single dishonest communicator about the location of an object that could be in one of two boxes. Participants were asked in which box the object was. These paradigms were adapted from Couillard and Woodward

(1999). To succeed in these tasks, children had to understand that the testimony was false and to use this understanding to infer the true location of the object.

We propose to call the kind of task we are using in Studies 2 and 3 false communication task (FCT) in order to highlight the commonalities between these and standard false belief tasks. In typical false belief tasks, participants know the true location of an object (that has been moved from one place to another). Their task is to understand that an individual who was not present when the object was moved would falsely believe it to be in its original location and would look for it there. In false communication tasks, participants know that a deceptive individual is claiming that an object is in one of two possible locations. Their task is to understand that this claim is false and to infer from this falsity the true location of the object. Both tasks involve understanding that the content of a representation—a belief, i.e. a mental representation in one case, a statement, i.e. a public representation in the other case—is false, and drawing true conclusions (about behavior in one case, about states of affairs in the other) from these false representations. Both standard false beliefs tasks and false communication tasks provide evidence that successful participants are capable of judging a representation to be false, thereby displaying some crucial epistemic understanding.

The task structure of false communication tasks avoids typical limitations of earlier paradigms. In one study (Lee & Cameron, 2000) for instance, young preschoolers had to learn after training to find out that a turtle was "in a basket" when a trickster said that it was not in the basket. In this case however, the trickster statement attracted attention on the correct hiding location. Across training, children may have come to know that the location alluded to by the "trickster" contained the turtle. This type of low level association was not possible in Studies 2 and 3. Moreover, in false communication tasks a mere affectively or cognitively driven disposition to avoid dishonest communicators would at best lead children to just ignore the deceptive claim, and therefore to answer at chance level. To do better, participants have to draw an epistemic inference from the falsity of the testimony to the true location of the object.

In Study 2, the communicator was directly described as a liar. In Study 3, on the other hand, a psychological inference from the mental dispositions of the communicator to the falsity of its testimony had to be performed before the epistemic inference.

5.1. Experiment 2a

5.1.1. Method

5.1.1.1. Participants. Fifteen 3-year-old children (M = 3;9, range 3;0 to 3;11) and 22 4-year-old children (M = 4;9, range 4;1 to 5;3) participated in the big liar test. Eighty four additional children were enrolled in control tests: thirty nine participated in the trust baseline assessment test (18 3-year-old, M = 3;6, range 3;1 to 4;0 and 21 4-year-olds, M = 4;7, range 4;1 to 5;0). Forty five children were enrolled in the disjunction test (20 3-year-old, M = 3;7, range 3;3 to 4;0 and 24 4-year-olds, M = 4;5, range 4;1 to 5;0).[1] All children were tested individually in their school.

5.1.1.2. Design. A single experimenter presented the task to participants in a room adjacent to their classroom. Children were tested on one of the three following conditions: trust baseline assessment, disjunction test or big liar test. To assess the consistency of children's answers, a subset of children (8 3-year-olds and 12 4-year-olds) participated in two big liar tests, without receiving feedback.

5.1.1.3. Procedure. The 'big liar' false communication task. The experimenter explained that he was going to hide a sweet in one of two boxes, the aim of the game being to find out where the sweet was. The experimenter then opened the boxes, asked children to turn round and loudly closed the boxes. The child was then invited to look again with the remark: "That's it! I have hidden the sweet!" A puppet frog was then introduced and made to look inside each of the boxes. A control question was asked to check whether children understood this part of the experiment: "Did the frog look into the boxes?" If children failed to answer properly, this phase of the procedure was repeated. If children failed again, the experimenter explicitly corrected them by saying: "No. the frog looked into the boxes". The experimenter then warned the child: "Now, the frog is going to talk to you, but be careful! The frog is a big liar! It always tells lies." The experimenter then checked whether children had registered the characterization by asking: "Is it a liar? Does it tell lies?" If children failed to answer, the description of the frog was repeated. No child required more than two repetitions to answer these control questions properly. The experimenter then announced that the frog was going to speak: he made the frog touch one of the boxes while saying in a distinctive voice, for instance, "The sweet is in the red box!". The participant was then asked the test question: "So, where is the sweet?". Whether the child selected the box indicated by the puppet or the other box was recorded. For those children who received two test trials, the experimenter said at the end of the first trial: "We are going to play this game again, but with two other boxes!" He then replaced the boxes used on the first trial by two boxes with a different shape and color. No indication was given to the child concerning whether he or she had answered correctly on the first trial.

Trust baseline assessment: This test was similar to the big liar test, except for the characterization phase. In this case, the puppet was not described as a liar, and the experimenter simply said: "Now, the frog is going to speak to you" without adding any particular warning.

Disjunction test: The disjunction test was identical to the trust baseline test, except that the puppet said: "The sweet is not in the red box". Since children knew that the sweet was either in the red box or in the green

[1] The slightly unequal sample sizes reported in the different conditions and studies result from the recruitment procedure. For each planned series of test, the experimenter enrolled all the available children in each participating school. Moreover, age groups were planned on the basis of grades, the exact age of children being collected *after* the test sessions, thus modifying the distribution in age groups.

box, they now had all the information needed to infer that it was in the green box. The aim of this test was to ascertain whether children in the age range we tested were able to perform this classical "or-elimination" deduction.

5.1.2. Results

5.1.2.1. Trust baseline assessment. Two 3-year-olds refused to complete the experiment. For the remaining sample ($n = 16$, $M = 3;6$, range 3;1 to 4;0, and $n = 21$, $M = 4;7$, range 4;1 to 5;0), the results for the trust baseline assessment task were straightforward: all children trusted the informant. This result calls for a methodological comment: there are two baselines for tests structured as the big liar test: First, chance level (50% of correct answers), which serves as a basis for evaluating whether the *majority* of children within an age group pass the test; Second the level of performance corresponding to a systematic trust in what the malevolent informant said (0% correct answers). Any significant deviation from this second level of performance indicates that a *significant subset* of a given age group passes the test. Both of these baselines are of relevance in assessing children's level of performance, and we used them both in Studies 2a, 2b, and 3.

5.1.2.2. Disjunction test. Three-year-olds and four-year-olds performed at ceiling on the disjunction test (respectively 100% and 92% of correct answers). These levels of performance were above chance ($p < 0.0001$ in both cases; binomial tests).

5.1.2.3. Big liar false communication task. Consistency of children's answers: Nineteen children out of the twenty consistently trusted or mistrusted the deceptive informant in both big liar tests. This pattern was different from what would be predicted by chance ($n = 20$; $p = 1.10^{-10}$; binomial test). The agreement between performance on the first and the second test was of 95%, (κ-coefficient = 0.9; 95% confidence interval from 0.71 to 1.09).

Results by age: Because only a subset of children was tested twice, subsequent analysis was performed on the results for the first liar test only. All 3-year-olds trusted the lying character, thus leading to a performance score of 0%, significantly below a chance score of 50% of correct answers ($p = 0.0001$; binomial test), and identical to the baseline trust level. Conversely, 4-year-olds gave 77% of correct answers. This result was significantly above that of 3-year-olds ($p = 2.10^{-6}$; Fisher Exact test), and above the chance level of 50% of correct answers ($p = 0.016$; binomial test). To further explore this sharp developmental trend, children were divided into four age groups, by six-month intervals: Young 3-year-olds ($n = 7$, $M = 3;3$, range 3;0 to 3;6), old 3-year-olds ($n = 8$, $M = 3;9$, range 3;7 to 3;11), young 4-year-olds ($n = 7$, $M = 4;4$, range 4;0 to 4;6), old 4-year-olds ($n = 15$, $M = 4;11$, range 4;6 to 5;3). The only significant increase from one age group to another was found between old 3-year-olds and young 4-year-olds: in a half year, children shifted from 0% to 86% correct answers ($p = 0.001$; Fisher Exact test).

5.2. Experiment 2b

Although some studies have demonstrated a rather refined understanding of the verb "to lie" among 3-year-olds (Siegal & Peterson, 1996, 1998), it is possible that younger children in Experiment 2a experienced comprehension difficulties in this particular testing context. It is also possible that they understood what the experimenter said, but did not believe that the puppet – handled by a benevolent experimenter – would actually deceive them. Additionally, since liars do not deceive all the time, younger children may have assumed that the "liar" was not going to lie during the game. Given however that the characterization of the puppet as a liar was part of a warning, the pragmatics of Study 2a heavily suggested that the informant was going to lie. Older preschoolers for example consistently interpreted this warning as an indication of the misleading aspect of the testimony. Still younger children may have expected a lower base rate for deception occurrences than older children, or may have differed in their pragmatic understanding of the warning. To control for these possibilities, we performed a repeated big liar test with feedback and debriefing. In this case, children's attention was repeatedly attracted on the discrepancy between what the misleading character said and the real state of affairs.

5.2.1. Method

5.2.1.1. Participants. Forty-nine children were recruited from two schools of a middle-size city in southwestern France. One 36-month-old and one 51-month-old child refused to complete the experiment. The remaining participants were divided into three age groups: 3-year-olds ($n = 13$, $M = 3;6$, range 3;0 to 3;11), 4-year-olds ($n = 23$, $M = 4;6$, range 4;1 to 5;11), and 5-year-olds ($n = 12$, $M = 5;3$, range 5;0 to 5;5).

5.2.1.2. Design. Children participated in two "big liar" tests without feedback and then received up to six tests sessions with training. Participants were tested in their school, in a quiet room adjacent to their classroom.

5.2.1.3. Procedure. Training for big liar test: Children participated in two repeated "big liar" trials without feedback as described in Experiment 2a. The only differences were that a "marble" was hidden instead of a "sweet". This part of the test provided the initial score of children. After the second trial, the experimenter provided feedback by opening the boxes and revealing the real location of the marble. Debriefing was provided by asking: "Which box did the frog tell you to choose? And where is the marble really?" If children failed on either of these control questions, they were corrected: "No, the frog told you to look in this box [showing the empty box], but the marble really was in this box [showing the box containing the marble]" and then the control questions were asked again. The control questions could be asked up to three times, and each time the experimenter corrected the child if he or she failed to answer properly. At the end of the procedure, the experimenter took the deceptive puppet and said: "the frog lied again!" Next, the child participated in repeated big liar tests, each time followed by feedback and debriefing. The procedure

was repeated until the child had passed at least two consecutive "liar tests". If this did not occur, the procedure ended after the eighth session of the liar test. Children received a length-of-learning score which amounted to the number of repeated trials they needed to pass two consecutive test trials.

5.2.2. Results

5.2.2.1. Initial score by age groups. As in Experiment 2a, children's performance was highly consistent. Only eight children out of 47 mistrusted the puppet on one test, while trusting it on the second ($p = 2.10^{-14}$; binomial test). This consistency of answers was present in each age group. 69% of 3-year-olds, 91% 4-year-olds and 83% of 5-year-olds kept using the same strategy on the two tests before feedback (either trusting or mistrusting the puppet). This level of consistency was above what could be predicted from chance for 4- and 5-year-olds ($p = 6.10^{-11}$ and $p = 0.03$; binomial tests). This tendency did not reach significance for 3-year-olds ($p = 0.26$, binomial test), most plausibly because of sample sizes. Because the tests of Study 2a and 2b were similar before feedback, a better measure of the consistency of 3-year-olds' answers could be assessed by summing results of Study 2a and 2b. This procedure revealed that the tendency for answering consistently was also above chance in 3-year-olds (80% of consistent answers, $p = 0.007$; binomial test).

Children's level performance was first assessed against chance (baseline of 50% of correct answers). A clear developmental pattern was found: 3-years-olds performed below chance level (15% of correct answers, $W_+ = 0$, $W_- = -45$, $p = 0.003$; Wilcoxon signed rank test). Four-year-olds' pattern of answers (35% of correct answers) did not differ significantly from 50% correct performance ($W_+ = 77$, $W_- = -154$, $p = 0.13$; Wilcoxon signed rank test). And 5-year-olds performed significantly above 50% correct performance (83% correct answers, $W_+ = 49.5$, $W_- = -5.5$, $p = 0.013$; Wilcoxon signed rank test). A Kruskal–Wallis test confirmed a main effect of age for children's initial scores ($KW = 14.59$, $p = 0.0007$). Subsequent Dunn multiple comparison tests indicated that the level of performance of 5-year-olds was significantly higher than the performance of 3- and 4-year-olds (respectively $p < 0.001$ and $p < 0.01$). Three- and four-year-old's level of performance did not differ significantly.

Since in the false communication task failing children have a tendency to consistently follow the advice of the puppet, a comparison with a systematic trust baseline (0% of correct answers) was also implemented.[2] Three-year-olds' performance on their first test essay was not different from the systematic trust baseline (1 success out of 13 children, $p = 1$; Fisher exact test). Four- and five-year-olds however performed above the systematic trust baseline (respectively eight successes out of 23 children, $p = 0.004$, and 11 successes out of 12 children, $p = 1.10^{-4}$; Fisher exact tests).

Comparing the performance of children on the first session of FC task revealed a significant sample effect: 4-year-olds enrolled in Experiment 2a performed better than children in Experiment 2b ($n = 45$, $p = 0.006$; Fisher Exact test).

5.2.2.2. Comparison of learning patterns. To compare children's length of training with the baseline of systematic trust, children categorized in two groups: participants who managed to reach two consecutive successes before the end of training, and participants who participated in all the training sessions. These distributions were assessed against a theoretical pattern of systematic trust (all children requiring all the training sessions). Ten 3-year-old out of 13 participated in six tests following feedback, a level of performance which did not differ from the baseline of systematic trust ($p = 0.22$; Fisher Exact test). Four- and five-year-olds length of training on the other hand were significantly below the baseline of systematic trust (respectively four children requiring full training out of 23, $p = 4.10^{-9}$ for 4-year-olds, and no children requiring full training out of 12, $p = 7.10^{-7}$, for 5-year-olds; Fisher Exact tests).

Three-year-olds required more training than 4- and 5-year-olds ($p = 0.0001$, in both cases; Fisher exact tests). The number of training sessions for 4- and 5-year-olds did not differ significantly ($p = 0.27$; Fisher Exact test).

5.2.2.3. Improvement by age: comparison of scores before and after training. The improvement for each age group was assessed by comparing the performance of children on the two initial tests with their score at the end of training. Note that this comparison overestimates improvement, because training was stopped when children succeeded on two consecutive trials. Children's improvement is represented on Fig. 1. Wilcoxon signed rank tests for matched pairs indicated that only 4-year-olds ($W_+ = 115$, $W_- = 0$, $p = 0.0005$) displayed a significant increase in performance from pre-training trials to the end of training phase. The increase in performance of 3- and 5-year-olds was not statistically significant (respectively $W_+ = 13$, $W_- = 2$, $p = 0.13$ for

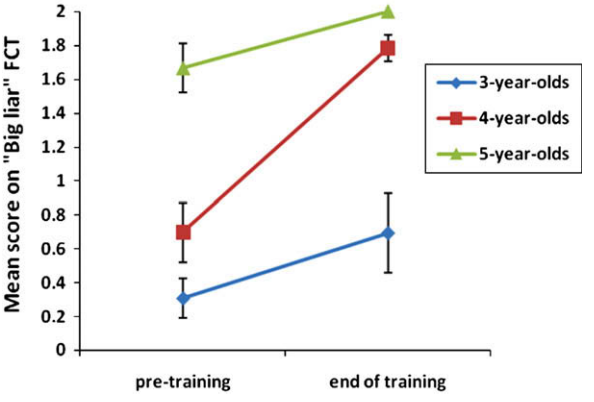

Fig. 1. Performance increase on the "big liar" false communication task after training by age group.

[2] Children's scores on their first test essay were assessed against this baseline using Fisher Exact tests. These particular comparisons and statistics were chosen because the assumption of symmetrical distribution around the median-required for Wilcoxon signed ranks tests- does not hold for comparisons with a theoretical median of correct answers.

3-year-olds and $W_+ = 6$, $W_- = 0$, $p = 0.1$, for 5-year-olds; Wilcoxon signed rank tests for matched pairs). Note however that 5-year-olds started with a relatively high score, leaving little room for improvement.

To assess the difference between training effects for the different age groups, an improvement score was computed by subtracting the initial score from the score at the end of training. This produced an index ranging from −2 to 2. A Kruskal–Wallis test indicated that the main effect of age on children's improvement score was significant ($KW = 7.12$, $p = 0.028$), although no post-hoc Dunn's comparisons between groups reached significance.

5.2.3. Discussion of Experiments 2a and 2b

In all three age groups, children took account of the content of the "big liar's" testimony. However, they clearly shifted from one strategy to another in the course of development. In Experiment 2a, all 3-year-olds consistently trusted the testimony of the "big liar". In Experiment 2b, the majority of 3-year-olds remained trusting even when they repeatedly experienced deceptive behavior. Older children on the other hand consistently mistrusted the misleading puppet, both before and after training.

Slight sample differences were observed: in Experiment 2a, the majority of 4-year-olds passed the big liar test without training, whereas in Experiment 2b, 4-year-olds reached ceiling only after training. Overall, however, the developmental pattern remained consistent: the youngest children accepted the testimony of the liar, older children were able to use feedback to improve their performance, and still older children were close to ceiling even before receiving feedback. Our results suggest that this developmental change takes place over a period of six months to one year that begins before the age of four.

The false communication task we employed had a specific aspect when compared to most lies encountered in real life. In most of genuine cases of deception, knowing that a lie is a lie falls quite short of telling you what the truth is. For example if John lies, "I am an astronaut", knowing that this is a lie does not tell you what his real occupation is for many alternatives are still possible. What you do learn is that John is not an astronaut. In the false communication task, being told, say, that "the marble is in the red box" is a lie also first tells you that the marble is not in the red box. But given that you know that the marble is either in the red or in the green box, you may come to the positive conclusion that the ball is in the green box by performing a disjunctive inference.

Could the need to perform a disjunctive inference in order to solve the task create an extra demand on participants leading one to underestimate the ability of the younger children to understand falsity? Results on the disjunction test provided strong evidence against such a hypothesis. Inferring the actual location of marble once told that it was not in a given box was extremely easy for all age groups. Three-year-olds in particular reached 100% of correct answers on the disjunction test.

More interestingly, given that the content of most lies is relatively uninformative (in ordinary contexts), young children might merely discard or ignore lies, rather than try to infer some truth from knowledge of their falsity. This conjecture might neatly explain why three-year-olds succeed when weighting information from two informants (e.g. in trust adjustment paradigms reviewed in Harris, 2007) while having difficulties when presented with a false communication task in which they have to infer the true state of affairs from the content of a single misleading testimony. However our results did not support this conjecture. In the false communication task, the only information given about the location of the marble is found in the lie of the puppet. Ignoring or discarding the lie would lead to answering at chance. If younger children had done so, they should have agreed 50% of the time with the misleading testifier (and selected the correct box 50% of the time). This pattern of answers was almost never observed. On repeated tests, individual performances were highly consistent in all age groups. Three-year-olds consistently trusted the informant before feedback. In Study 2b, even after extensive training, most 3-year-olds still trusted the informant. This pattern of answers leaves two possibilities: either the younger children failed to recognize that the liar's statement was false, or they did recognize this but did not have the executive capacity to translate this recognition into appropriate behavior.

The "big liar" FCT used in Experiments 2a and 2b is an adequate test of children's ability to understand the falsity of lies and to draw epistemic inferences from it. On the other hand, this kind of test does not, by itself, show a grasp of the intentional aspect of deception. Passing it is compatible with a partial understanding of lying as "saying something false" without the deceptive intention of the informant being taken into account. This is not just a theoretical possibility: some experiments suggest that before 6 to 8 years of age, children have a tendency to describe false statement as "lies" regardless of the intentions of the speaker (Peterson, Peterson, & Seeto, 1983; Strichartz & Burton, 1990; Wimmer, Gruber, & Perner, 1984).

6. Study 3: Understanding the intention to deceive

We developed and used in Study 3 a FCT designed to test participants' ability to use mindreading understanding of deception in filtering information. The structure of the task remained similar to that of the big liar FC task used in Study 2, thereby precluding the use of mere moral preferences for passing the test. This time however, the communicator was not described as a liar, but merely as a very mean character who did not want the child to find a sweet hidden in one of two boxes. Children had therefore to infer that the communicator would be trying to mislead them by intentionally giving them false information before they could infer the true location of the sweet from the falsity of the communicator's testimony. Even though the word 'lying' was not used in the test, children who passed the test thereby showed not only an epistemic but also a mindreading understanding of lying.

To investigate the consistency of children's strategies, they were presented with a repeated test without feedback. A post-test interview was administered in order to clarify whether children who had not been vigilant were

nevertheless able to characterize the behavior of the naughty character as lying.

6.1. Experiment 3

6.1.1. Method

6.1.1.1. Participants. Eighteen 4-year-olds (M = 4;5, range 3;11 to 4;10), 30 5-year-olds (M = 5;6, range 5;1 to 6;0), and 13 6-year-olds (M = 6;3, range 6;1 to 6;5) participated in the study. All were recruited from two schools in southern France.

6.1.1.2. Design. Children were tested at their school by a single experimenter. They were presented with two false communication test trials, and a post-test interview.

6.1.1.3. Procedure. The "mean" false communication task. The testing procedure was similar to the big liar test except for the characterization phase and the post-test interview. In the characterization phase, the experimenter said: "Be careful! The frog is very mean! It does not want you to find the sweet!" Control questions were: "Is it mean?"; "Does it want you to find the sweet?" As in Study 2, descriptions of the frog were repeated if children failed. Children never required more than two repetitions of the description to answer these control questions properly. At the end of the second testimony trial, the experimenter gave children feedback on the location of the sweet. Only the experimenter looked into the boxes (that were actually empty) and commented on their contents. When opening the box that the deceptive informant had indicated, the experimenter said: "It's not there!" and, when looking into the other container, he said: "It's in here". The experimenter always talked first about the box that the child had selected. Children were then asked a memory question: "Which box did the frog tell you to choose?" and a reality question: "and where was the sweet really?" Children were then presented with two open-ended questions in which they had an opportunity to spontaneously express the fact that the puppet had lied or tricked them:

- Question 1: "The frog told you to open this box [showing the "empty" box], but in fact the sweet was in that one [showing the other box]. How is that possible?"
- Question 2: "The frog knew that the sweet was in this box [showing the correct box], but he told you to choose the other one. How do you call what frog did?"

If the child did not spontaneously describe the behavior of the character as lying, the experimenter asked a forced-choice question: Half the children were asked "Did the frog lie or did he make a mistake?" and the other half was asked: "Did frog make a mistake or did he lie?" to control for potential order effects. If children answered any of these three questions correctly, they were recorded as passing the lie identification test.

6.1.2. Results

6.1.2.1. Consistency of answers. Overall, the distribution of scores revealed two main strategies. Most children either always trusted or always mistrusted the malevolent character. The consistency in children's strategies across the two FC trials was 83.61%, a good level of agreement as indexed by kappa's test (κ-coefficient = 0.66; 95% confidence interval from 0.48 to 0.85). The proportion of children performing inconsistently (i.e. trusting the puppet on one occasion and mistrusting the puppet on the other) was below chance for 4-year-olds (6% of children, n = 18, p = 0.0001, binomial test), 5-year-olds (26% of children, n = 30, p = 0.005; binomial test) and 6-year-olds (13% of children; n = 16; p = 0.004; binomial test).

6.1.2.2. Results by age. Testimony test. The performance of children on the first and second FC trial did not differ significantly (p = 0.34; McNemar's test). Thus, scores for the two trials were combined into a single measure of performance. Four-, five- and six-year-olds mistrusted the misleading informant respectively in 30%, 58% and 92% of cases. Only 6-year-olds' performance was significantly above 50% of correct answers (W_+ = 66, W_- = 0, p = 0.001, for 6-year-olds; Wilcoxon signed rank tests). Four- and five-year-olds' performance was not significantly different from a baseline of 50% of correct answers (respectively W_+ = 45, W_- = −108, p = 0.095, for 4-year-olds and W_+ = 168, W_- = −108, p = 0.30, for 5-year-olds; Wilcoxon signed rank tests). A Kruskal–Wallis test revealed a main effect of age on performance (KW = 13.5, p = 0.0012). Dunn's multiple comparisons tests indicated that only the difference between 4- and 6-year-olds was significant (Mean rank difference = −21.7, $p < 0.0001$).

To assess children's performance against the systematic trust baseline, children's performance on their first test essay was again considered. All age groups performed above the systematic trust baseline on their first essay (respectively six successes out of 18 children, p = 0.02, for 4-year-olds, 19 successes out of 30 children, $p = 5.10^{-8}$, for 4-year-olds, and 12 successes out of 13 children, $p = 3.10^{-5}$, for 6-year-olds; Fisher exact tests).

Interview: Memory and explicit characterization test. Results on the interview questions are reported in Table 1. Four-year-olds did no better than chance on the memory test (61% correct answers) and when deciding whether the misleading informant told a lie or made a mistake (50% correct answers). Five- and six-year-olds reached ceiling on the memory test and when explicitly characterizing the behavior of the misleading puppet as a lie. The increase

Table 1
Percentage of correct answers on the interview questions (memory test and explicit characterization test) as a function of age.

	Four-year-olds n = 18	Five-year-olds n = 30	Six-year-olds n = 13
Memory test (binomial test, p-value)	61% (0.48)	90% (<0.0001)	100% (<0.0001)
Explicit characterization (binomial test, p-value)	50% (1)	83% (0.0003)	92% (0.003)

in performance from four to five years was significant both on the memory test ($n = 48$, $p = 0.002$; Fisher Exact test) and on the lie identification question ($n = 48$, $p = 0.02$; Fisher Exact test).

6.1.3. Discussion of Experiment 3

The findings of Study 3 were in agreement with those of Experiments 2a and 2b. The consistency of children's strategies shows that they did not simply ignore the testimony of the malevolent character: they took it into account, either accepting it or rejecting it. Ignoring what the informant said would have lead to answer at chance (leading to a score of 50% of correct answers). Four-year-olds' performance (30% of correct answers) did not differ significantly from this level of performance. However, considering the consistency of children's answers reveals that among 4-year-olds, a subset of children was consistently vigilant (always disagreeing with the misleading testimony), and an other subset consistently agreed with the misleading testifier. The pattern of agreeing half of the time with the testimony of the puppet was rarely observed (e.g. for 4-year-olds: 6% of children, $n = 18$, $p = 0.0001$; binomial test).

Subsequently, although the majority of 4-year-olds failed, a subset was able to pass the task. By this age, children start to display fully-fledged capacity for epistemic vigilance towards deception. They were able to infer the intent to deceive from a malevolent disposition and to take advantage of misleading information in order to infer the truth from it.

The interview questions revealed two surprising results: On the memory question, a subset of children maintained that the misleading informant had told them the correct location of the sweet. This "memory failure" is particularly surprising because most of the children who experienced it (6 out of 8) actually trusted the informant before answering the memory question. For this particular test children had not to predict the behavior of the testifier, but merely to acknowledge that he provided a false testimony. Rather, in this case, many younger children were the victims of the deception, but did not seem to interpret it as such. They erased the misleading testimony from their memory. A similar effect was suggested as potential interpretation for difficulty to accurately remember what an informant suggested in a false belief task context (Robinson, Mitchell, & Nye, 1995), in another context with children trusting informants rather than their own perception in an Asch consensus paradigm (Corriveau & Harris, personal communication), and in a related unpublished study by the first author. Maybe these children were so trusting that the tendency to revise their memory of the informants' testimony was stronger than the tendency to revise their belief in the informant's trustworthiness. Another possibility is that they did not keep track of what the informant had said. Once they knew for sure where the sweet was, they inferred the claim made by the misleading informant from their own knowledge. Whatever the cognitive explanation for this behavior, children who failed the memory test placed themselves at risk of being deceived again by the misleading puppet.

Another surprising result of this Study was that younger children did not accurately describe the behavior of the mean character as a lie, whereas several studies have provided evidence that lie identification is possible for children as young as three years of age (Gilli, Marchetti, Siegal, & Peterson, 2001; Siegal & Peterson 1996, 1998). A tentative interpretation of this difference is that in the studies by Siegal and his colleagues, children were ascribed the point of view of the deceiver, whereas in the current Study, children had to take a perspective different from their own in order to characterize the behavior of the naughty puppet as a lie.

7. General discussion

Children engage in cooperative forms of communication at the beginning of their second year of life if not earlier. For instance, preverbal children as young as 12 months of age help people find the objects they are looking for by pointing (Liszkowski, Carpenter, Striano, & Tomasello, 2006) and seem to be motivated to provide people with relevant information (Liszkowski, Carpenter, & Tomasello, 2007). These young children are typically as trusting in matters of communication as they are in all matters of nurturing, and how could it be otherwise? If the adults who are taking care of them are not trustworthy, there is very little they can do about it. As they grow older and become more autonomous, children communicate more and more with other children and adults who may not have their best interest at heart. Exerting some vigilance becomes both more appropriate and more feasible.

Here we have looked at the particular case of the development of epistemic vigilance towards deception. Our results suggest that such vigilance has three components: a moral, an epistemic and a mindreading one. The fact that these three components have each a distinct developmental trajectory suggests that they are not just analytically distinct but also that they are subserved by distinct mental mechanisms.

7.1. The moral component

Study 1 showed that children as young as 3-year-olds favor testimonies provided by benevolent rather than by malevolent informants. The ability to distinguish benevolent from malevolent agents and to prefer the formers probably develops early in infancy, as shown by studies in which young infants display a preference for helping attitudes (e.g. Wynn, 2007). One might speculate that this is a basic self-protection mechanism likely to be evolutionarily ancient. Indeed, there is evidence that it is found in other species (e.g. Bates et al., 2007 for data on elephant categorization which may be founded on the basis of benevolence). What is specific to our results is the demonstration that this general preference for benevolent agents leads to a specific preference for their testimony over that of malevolent agents. In our experiments, both the nice and the mean puppets informed the child of the identity of the object in the box in the same purportedly helpful manner. Still, children drew on the information they had

previously acquired regarding the benevolence or malevolence of the two characters and used it in deciding which one to believe. The depth of moral understanding underpinning this behavior remains to be ascertained. Children may have just treated the malevolent puppet as dangerous and to be avoided, or they may have seen it not just as dangerous but also as blameworthy. Either way, these evaluations have been sufficient to drive children's selection of testimony. An affective preference for benevolent agents is not particularly geared to filtering communicated information, but it can be put in the service of this function.

These findings dovetail with those of recent studies focused on epistemic vigilance not towards malevolence but towards incompetence that have nevertheless shown positive effects of familiarity (Corriveau & Harris., 2009; Harris et al., in press), or attachment (Corriveau et al., 2009) on children's level of trust. In these studies, children seem to trust familiar informants not just because they have proven reliable in multiple past encounters but also for affective reasons. This suggests a possible partial reinterpretation of previous work on children's selective trust (reviewed in Harris, 2007). Even when selecting the testimony of informants who have previously been accurate, young children may, at least in part, be driven by moral and affective preferences. There is evidence of an early affective preference for agents who get social conventions "right", in particular in the domain of vocabulary use (Rakoczy, Warneken, & Tomasello, 2008; Sabbagh & Baldwin, 2001). Reliance on this type of global good/bad assessment might in particular explain why in studies of informant accuracy, 3-year-old children do not adjust their trust to the frequency of accurate testimonies, but rather discriminate between informants who have always been accurate, and those who have been inaccurate at least once (Pasquini, Corriveau, Koenig, & Harris, 2007). A single error may be sufficient to provoke a negative affective reaction that, in turn, yields epistemic mistrust. This might also explain why children sometimes mistrust informants who were inaccurate but for a good reason, e.g. because they were blindfolded (Nurmsoo & Robinson, 2008).

Nevertheless, we do not believe that results on precocious epistemic trust adjustment are likely to be entirely reduced to affective preferences for positive character traits. While affective preferences may drive part of the effects in studies contrasting informants' level of accuracy, they are for example unlikely to explain young children's preferences for better informed testifiers (Nurmsoo & Robinson, 2009; Robinson & Nurmsoo, 2009; Robinson & Whitcombe, 2003). The ability to assess the relative informative value of two contrasting sources of information is thus probably in place around the age of four, possibly before. Younger children may for example tag the information provided by the misleading informant as somehow defective (rather than as false). This would allow them to succeed in selecting the more appropriate message when two informants are contrasted. Whether this type of assessment—rather than mere affective preferences—is at the basis of young children's good performance in the benevolence paradigm of Study 1 deserves further investigations.

7.2. A methodological comment on false communication tasks

When there is only one informant, holding his or her contribution as somehow defective does not help select the right box. To resist lies in such cases rests on the ability to recognize categorically that the information communicated by a single mistaken or misleading informant is probably false. Studies 2 and 3 directly addressed this issue with the use of false communication tasks. In building these tests, we were guided by considerations comparable to those that drove the design of false belief tasks in the Theory of Mind literature (e.g. see Bennett, 1978; Dennett, 1978; Harman, 1978). The tests were built to ensure that children could not answer properly without treating the content of the testimony as false. As a result, success on a false communication task reliably indicates that the participant understands that communicated information may be false. Failure on the task by young children however, is less diagnostic and requires thorough methodological discussions (for similar arguments in the domain of false belief tasks, see Bloom & German, 2000).

Having to treat communicated information as false, as they had to do in Studies 2 and 3, may be challenging for younger children. They may not possess the cognitive resources for epistemic categorization (into 'true' and 'false') and for epistemic inference (from [P is false] to [not P]). The study of strategic deception (Peskin, 1992; Ruffman, Olson, Ash, & Keenan, 1993; Russell, Mauthner, Sharpe, & Tidswell, 1991; Sodian & Frith, 1992) and false sign tasks (Bowler, Briskman, Gurvidi, & Fornells-Ambrojo, 2005; Leekam, Perner, Healey, & Sewell, 2008; Sabbagh, Moses, & Shiverick, 2006) provides evidence that this could be so. Young children may develop the capacity to process the epistemic status of representations at about the time they become able to pass the standard false belief tasks.

In Studies 2 and 3, young children also had to infer a tendency to produce misinformation (e.g. from being a "liar" or being "mean" to telling a lie) from a negative disposition. This requirement may have challenged younger participant for yet another reason. Children sometimes evidence a positivity bias in making attribution. They have a tendency to make more stable attributions of internal traits to others on the basis of positive events than on the basis of negative ones (Boseovski & Lee, 2006; Heyman & Giles, 2004; Rholes & Ruble, 1984; although for counter-evidence see Aloise, 1993). Preschoolers also have sometimes a stronger tendency (compared to older children and adults) to expect positive changes in others' negative behavior (Lockhart, Chang, & Story, 2002; Lockhart, Nakashima, Inagaki, & Keil, 2008; Solomon, Johnson, Zaitchik, & Carey, 1996). Such a positivity bias may impact on children's selection of partners for cooperative enterprises (see e.g. teammate preferences for academic competition in Droege & Stipek, 1993). Since younger children are more dependent on caregivers and have less choice in the selection of partners to interact with, they may be less willing and able to categorize people as malevolent. In particular, young children may be unable to catch people who are covertly uncooperative. For example, young preschoolers

have been found to have difficulty understanding that a mean character pretending to be someone nice would play mean tricks (Peskin, 1996). This would explain younger children difficulties when "liars" or "mean" informants seem to engage in a cooperative sharing of information (by communicating).

Both insufficient epistemic competence and a positivity bias may be relevant to explaining why younger children have trouble comprehending the falsity of lies and deceptive intents.

7.3. The epistemic component

Success on false communication tasks can, on the other hand, be taken as a reliable indicator of many important capacities underpinning vigilance towards deception. To pass the "big liar" FCT of Study 2, children needed to:

1. Be able to resist the suggestion of the misleading testimony and to select the other box, which may have posed high demands in terms of executive functioning skills.
2. Associate the puppet characteristics to a tendency to produce misinformation, thus overcoming a potential positivity bias.
3. Make the epistemic judgment that what the puppet said is false. A child could have a half-understood notion of what a lie is, understanding for instance that a lie is morally wrong and not to be accepted, without properly grasping the notion of falsity.
4. Draw appropriate inferences from the content of the false testimony. Children had first to draw the epistemic inference from, e.g. "'the marble is in the red box' is false" to "the marble is not in the red box," and then to draw the disjunctive inference from, e.g. "the marble is either in the red box or in the green box" and "the marble is not in the red box" to "the marble is in the green box."

Our result shows that by the age of four years, children are able to implement all these steps. In particular, children passing the FCT evidence the capacity to assess the truth or falsity of communicated messages, and to make use of this assessment in allocating their trust and in drawing inferences.

7.4. The mindreading component

Both the epistemic ability to recognize the falsity of a statement and the mindreading ability to understand a deceptive informant's motive in producing a false statement are needed for fully-fledged epistemic vigilance. The epistemic ability could in principle exist without the mindreading ability (but not the other way around). Our Study 3 showed that this is not just a theoretical possibility. When they had to infer the falsity of a statement from an understanding of the deceptive intent of the informant, most 4-year-olds failed, and so did almost half of the 5-year-olds. More than 90% of 6-year-olds on the other hand were successful.

8. Assessing the developmental course of the model

Overall, the developmental pattern we observed differentiated three abilities relevant to children's epistemic vigilance:

(1) A moral/affective ability to prefer the testimony of a nice informant over that of a mean one, already in place at the age of three.
(2) An epistemic ability to recognize the falsity of lies on the basis of testifiers' dispositions, evidenced around the age of four.
(3) A mindreading ability to understand that an agent may intend to misinform his audience and do so by producing a lie, observed to develop between four and six years of age.

Our participants were all children from rural or small town schools in the South of France, living—this is an impressionistic comment—in a relatively friendly and trusting atmosphere. It is possible that, in different cultural contexts, mistrust is more encouraged or on the contrary discouraged, causing the developmental pattern we observed to be somewhat speeded up or slowed down. Economic conditions, gender, and position among siblings might also make a difference. All this would deserve investigation. It would be quite surprising, however, if, in another cultural context, the overall pattern we observed was altogether absent or wholly or partially reversed.

Acknowledgements

This research forms part of M. Mascaro doctoral thesis. It was supported by a grant from the Direction Générale de l'Armement to the first author and by the Center for the Study of the Mind in Nature (University of Oslo). The authors are indebted to Nicolas Baumard, Nicolas Claidière, Fabrice Clément, Coralie Chevallier, Maria Fusaro, Paul Harris, Christophe Heintz, Hugo Mercier, Olivier Morin, Gloria Origgi, Guy Politzer and Deirdre Wilson for invaluable inputs at different stages of this research. The authors also wish to thank Rebecca Gomez and three anonymous reviewers who suggested many stylistic, methodological and conceptual improvements in the course of the submission process. Last but not least, our warmest thanks are expressed to the teachers, parents and children of the 16 schools which participated in the project.

References

Adler, J. E. (1997). Lying, deceiving, or falsely implicating. *Journal of Philosophy, 94*, 435–452.
Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attributional biases. *Journal of Experimental Child Psychology, 55*, 177–193.
Alvarez, J. M., Roger, D. M., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children behavioral predictions. *Child Development, 72*, 1409–1425.
Bates, L. A., Sayialel, K. N., Njiraini, N., Moss, C. J., Poole, J. H., & Byrne, R. W. (2007). Elephants classify human ethnic groups by odor and garment color. *Current Biology, 17*, 1938–1942.
Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 4*, 557–560.

Bergstrom, B., Moehlmann, B., & Boyer, P. (2006). Extending the testimony problem: Evaluating the truth, scope, and source of cultural information. *Child Development, 77*, 531–538.

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1118–1134.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the standard false belief task as a test of theory of mind. *Cognition, 77*, B25–B31.

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behvioral predictions. *Developmental Psychology, 42*, 500–513.

Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development, 6*, 259–283.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an implicit theory of mind. *Developmental Psychology, 18*, 906–921.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development, 6*, 53–75.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.

Clément, F., Koenig, M. A., & Harris, P. L. (2004). The ontogenesis of trust. *Mind and Language, 19*, 360–379.

Corriveau, K. H., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science, 12*, 426–437.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., et al. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80*, 750–761.

Couillard, N. L., & Woodward, A. B. (1999). Children's comprehension of deceptive points. *British Journal of Developmental Psychology, 17*, 515–521.

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

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568–570.

Droege, K., & Stipek, D. J. (1993). Children's use of dispositions to predict classmates' behavior. *Developmental Psychology, 29*, 646–654.

Dunn, J. (1991). Understanding others: Evidence from naturalistic studies of children. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 51–61). Oxford: Basil Blackwell.

Eskritt, M., Whalen, J., & Lee, K. (2008). Young children's recognize violations of the Gricean maxims. *British Journal of Developmental Psychology, 26*, 435–443.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*, 77–83.

Freire, A., Eskritt, M., & Lee, K. (2004). Are eyes windows to a deceiver's soul? Children's use of another's eye gaze cues in a deceptive situation. *Developmental Psychology, 40*, 1093–1104.

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science, 11*, 771–777.

Gee, C. L., & Heyman, G. D. (2007). Children's evaluation of other people's self-descriptions. *Social Development, 16*, 800–810.

Gilles, J. W., & Heyman, G. D. (2005). Preschoolers use trait-relevant information to evaluate the appropriateness of an aggressive response. *Aggressive Behavior, 31*, 498–509.

Gilli, G., Marchetti, A., Siegal, M., & Peterson, C. (2001). incipient ability to distinguish mistakes from lies: An Italian investigation. *International Journal of Behavioral Development, 25*, 88–92.

Hala, S., Chandler, M. J., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of 3-year-olds' understanding of false belief. *Child Development, 62*, 83–97.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–560.

Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences, 4*, 576–577.

Harris, P. L. (2007). Trust. *Developmental Science, 10*, 135–138.

Harris, P. L., Pasquini, E., Corriveau, K., Koenig, M., & Clement, F. (in press). In J. Proust, J. Dokic, & E. Pacherie (Eds.), *From metacognition to self-awareness*. Oxford, UK: Oxford University Press.

Heyman, G. D., Fu, G., & Lee, K. (2007). Evaluating claims people make about themselves: The development of skepticism. *Child Development, 78*, 367–375.

Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development, 74*, 516–534.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inference. *Child Development, 70*, 604–619.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology, 77*, 1–19.

Heyman, G. D., & Giles, J. W. (2004). *Valence effects in reasoning about evaluative traits*. Merrill.

Heyman, G. D., & Legare, C. H. (2005). Children's evaluation of sources of information about traits. *Developmental Psychology, 41*, 636–647.

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-years-old sentivity to cues to credibility. *Journal of Cognition and Development, 8*, 263–283.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use reliability over age when learning new words. *Psychological Science, 17*, 757–758.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology* (pp. 380–402). Sunderland, MA: Sinauer Associates.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14*, 402–408.

Lampinen, J. M., & Smith, V. L. (1995). Incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. *Journal of Child Psychology, 80*, 621–627.

Lee, K., & Cameron, C. A. (2000). Extracting truth information from lies: The emergence of representation-expression distinction in preschool children. *Merrill Palmer Quarterly, 40*, 1–20.

Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology, 26*, 485–497.

Lewis, M., Stranger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology, 25*, 439–443.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development, 7*, 173–187.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). *Developmental Science, 10*(2), F1–F7.

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development, 78*, 1543–1558.

Lockhart, K. L., Chang, B., & Story, T. (2002). Children's beliefs about the stability of traits: Protective optimism? *Child Development, 73*, 1408–1430.

Lockhart, K. L., Nakashima, N., Inagaki, K., & Keil, F. C. (2008). From ugly duckling to swan? Japanese and American beliefs about the stability and origins of traits. *Cognitive Development, 23*, 155–179.

Matsui, T., Miura, Y., & McCagg, P. (2006). In *Proceedings of the 28th annual cognitive science society* (pp. 1789–1794).

Mills, C. M., & Keil, F. C. (2005). The development of cynicism. *Psychological Science, 16*, 385–390.

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false belief tasks. *British Journal of Developmental Psychology, 18*, 297–317.

Nurmsoo, E., & Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well or poorly-informed: When past errors can be excused. *Child Development, 80*, 23–27.

Nurmsoo, E., & Robinson, E. J. (2008). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science, 11*, 905–911.

Pasquini, E. S., Corriveau, K. H., Koenig, M. A., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Science, 43*(5), 1216–1226.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Peskin, J. (1992). Ruse and representations: On children's ability to conceal information. *Develomental Psychology, 28*, 84–89.

Peskin, J. (1996). Guile:children s understanding of narratives in which the purpose of pretense is deception. *Child Development, 67*, 1735–1751.

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development, 54*, 1529–1535.
Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*, 561–568.
Premack, D., & Premack, A. J. (1997). Infants attribute value to the goal directed actions of self-propelled objects. *Journal of Cognitive Neuroscience, 9*, 848–856.
Rakoczy, H., Warneken, F., & Tomasello, M. (in press). Young children's selective learning of rules game from reliable and unreliable models. *Developmental psychology*.
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*(3), 875–881.
Rholes, W., & Ruble, D. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*, 550–560.
Robinson, E. J., Champion, H., & Mitchell, P. (1998). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology, 35*, 535–546.
Robinson, E. J., & Nurmsoo, E. (2009). When do children learn from unreliable speakers? *Cognitive Development, 24*, 16–22.
Robinson, E. J., Mitchell, P., & Nye, R. (1995). Young children's treating of utterances as unreliable sources of knowledge. *Journal of Child Language, 22*, 663–685.
Robinson, E. J., & Whitcombe, E. L. (2003). Children s suggestibility in relation to their understanding about sources of knowledge. *Child Development, 74*, 48–62.
Ruffman, T., Olson, D. R., Ash, T., & Keenan, T. (1993). The ABC's of deception: Do young children understand deception in the same way as adults? *Developmental Psychology, 38*, 74–87.
Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331–349.
Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between theory of mind and semantic development. *Child Development, 72*, 1054–1070.
Sabbagh, M. A., Moses, J. L., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development, 77*, 1034–1049.
Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers? *Journal of Child Language, 30*, 905–924.
Scofield, J., & Behrend, D. A. (2008). Words from reliable and unreliable speakers. *Cognitive Development, 23*, 278–290.
Searcy, W. A., & Novicky, S. (2006). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton, New Jersey: Princeton University Press.
Shultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology, 17*, 465–471.
Siegal, M., & Peterson, C. C. (1996). Breaking the mold: A fresh look at children understanding of questions about liers and mistakes. *Developmental Psychology, 32*, 322–344.
Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34*, 332–341.
Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9*, 173–188.
Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 591–605.
Solomon, G. E., Johnson, S. C., Zaitchik, D., & Carey, S. (1996). Like father, like son: Young children's understanding of how and why children resemble their parents. *Child Development, 67*, 151–171.
Sperber, D. (1994). Understanding verbal understanding. In Jean Khalfa (Ed.), *What is intelligence?* (pp 179–198). Cambridge University Press.
Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford, UK: Oxford University Press.
Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics, 29*, 401–413.
Strichartz, A. F., & Burton, R. V. (1990). Lies and truth: A study of the development of the concept. *Child Development, 71*, 211–220.
Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology, 13*, 191–204.
Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development, 26*, 436–444.
Talwar, V., Murphy, S. M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31*, 1–11.
Welch-Ross, M. K. (1999). Interviewer knowledge and preschoolers' reasoning about knowledge states moderate suggestibility. *Cognitive Development, 14*, 423–442.
Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to the report the source of their belief. *Cognitive Development, 15*, 329–346.
Wilson, A. E., Smith, M. D., & Ross, H. S. (2003). The nature and effects of young children's lies. *Social Development, 12*, 21–45.
Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism-moral subjectivism. *Journal of Experimental Psychology, 1*, 1–30.
Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9*, 257–270.
Wojciszke, B. (2005). Affective concomitants of information on morality and competence. *European Psychologist, 10*, 60–70.
Wynn, K. (2007). Some innate foundations of social and moral cognition. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Foundations and the future*. Oxford: Oxford University Press.