

Why should I trust you? Investigating young children's spontaneous mistrust in potential
deceivers

Roman Stengelin^{1, 2, *}, Sebastian Grueneisen¹, Michael Tomasello^{1, 3}

¹Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Deutscher Platz 6
04103 Leipzig
Germany

²Leipzig Research Center for Early Child Development
Jahnallee 59
04109 Leipzig
Germany

³Duke University
417 Chapel Dr.
Durham, NC 27708
USA

* Corresponding author

E-mail: roman.stengelin@uni-leipzig.de
Jahnallee 59
04109 Leipzig,
Germany

##2018 Cognitive Development; doi: <https://doi.org/10.1016/j.cogdev.2018.08.006>##

Why should I trust you? Investigating young children's spontaneous mistrust in potential deceivers

Abstract

Children must learn not to trust everyone to avoid being taken advantage of. In the current study, 5- and 7-year-old children were paired with a partner whose incentives were either congruent (cooperative condition) or conflicting (competitive condition) with theirs. Children of both ages were more likely to mistrust information spontaneously provided by the competitive than the cooperative partner, showing a capacity for detecting contextual effects on incentives. However, a high proportion of children, even at age 7, initially trusted the competitive partner. After being misled once, almost all children mistrusted the partner on a second trial irrespective of the partner's incentives. These results demonstrate that while even school age children are mostly trusting, they are only beginning to spontaneously consider other's incentives when interpreting the truthfulness of their utterances. However, after receiving false information only once they immediately switch to an untrusting attitude.

Keywords:

- Epistemic Vigilance
- Selective Trust
- Cooperation
- Competition
- Deception

1. Introduction

Children heavily depend on socially transmitted information (e.g. when learning what food to eat or to avoid, what rules to follow, which animals might be dangerous, etc.). By taking advice from others they acquire knowledge and skills they could never learn by relying on personal experiences alone (Henrich, 2015). It is therefore not surprising that young children have a strong tendency to trust information provided by others (Jaswal, Croft, Setia, & Cole, 2010). However, while conferring clear advantages in most cases, trusting other's advice blindly entails a risk – such information could be false or even intentionally deceptive. Children thus have to learn to be vigilant when interpreting other's utterances and identify individuals who are both willing and able to offer truthful information (Sperber et al., 2010).

Indeed, children are highly selective when deciding whom to trust. For instance, from age 3 children selectively trust previously accurate informants over inaccurate ones (Jaswal & Neely, 2006; Pasquini, Corriveau, Koenig, & Harris, 2007; Corriveau & Harris, 2009). Around the same age, they also start to preferably trust informants who previously adopted majority views than informants who dissented from a consensus (Fusaro & Harris, 2008; Corriveau, Fusaro, & Harris, 2009) and on whether past false statements of informants can be excused by situational factors, such as lack of perceptual access (Nurmsoo & Robinson, 2009a; but see Nurmsoo & Robinson, 2009b). While these studies show that children's abilities to evaluate informants' competence at providing truthful information develop early only recently have studies started looking at whether children also consider others' intentions when deciding whether or not to trust their utterances.

How children deal with utterances provided by potentially dishonest informants is far from being fully understood. Informants may intentionally provide misleading information for several reasons. For instance, they may deceive others because they are mean by disposition. Accordingly, children have been shown to mistrust individuals who were introduced as being mean or who were labeled a liar and this tendency increases between age three to six (Mascaro & Sperber, 2009). Moreover, preschoolers tend to trust informants labeled as nice more than informants who they were told have relevant expertise indicating that others' alleged character dispositions play an important role in children's trust decisions (Landrum, Mills, Johnston, 2013). Another source for inferring other's intentions is their prior behavior: Five-year-olds tend

to mistrust hints by informants who they have previously seen showing positive affect when misleading others (Vanderbilt, Liu, & Heyman, 2011; Heyman, Sritanyaratana, & Vanderbilt, 2014). Observing an informant who is trying to trick others also decreases children's trust even if this information leads to positive outcomes (Liu, Vanderbilt, & Heyman, 2013). These studies thus indicate that between age 4 to 5, children already consider others' intentions when deciding whether or not to trust them (see also Shafto, Eaves, Navarro, & Perfors, 2012). However, 5- and 6-year-olds give greater weight to past accuracy than to observed intent (Liu, Vanderbilt, & Heyman, 2013) suggesting that children's abilities for intent-based trust are still somewhat fragile at this age.

Much less attention has been paid to another way of finding out about others' intentions, and as a consequence, their trustworthiness, namely, to consider their incentives in a given situation. The ability to do this is of particular importance since, when interacting with others in everyday life, we do not necessarily have clear information about their character dispositions or their past behaviors in similar situations. Individuals usually have to decide spontaneously and without prior warning whether or not to trust others. Deceivers tend to conceal their intentions, so that decisions on whether to trust informants often have to be inferred from the social context alone. For example, a salesperson may overstate the value of his/her goods in order to sell them at a high price and potential buyers have to be vigilant about the other's motives in order to avoid being tricked. We should thus preferentially trust others who have incentives for cooperation (i.e., incentives that are congruous with our own) compared to individuals who have incentives for competition (i.e., incentives that are in conflict with our own). To our knowledge, only two studies have addressed the question whether children consider an informant's intentions as a function of their incentives and without receiving prior information about the informant's disposition or past accuracy.

In one study, 3-9-year-olds were confronted with a "Sender-Receiver Game" in which children were asked to search for a hidden reward in one of two boxes (Sher, Koenig, & Rustichini, 2014). An adult experimenter (E1) who knew the location of the reward was asked by a second experimenter to provide a hint about its location. If children picked the correct box they could keep the reward and E1 received nothing. E1 only got the reward if children picked the wrong box such that providing truthful information was contradicting E1's self-interest.

Children's mistrust in E1's hints increased with age. Around half of the 5-year-olds were initially skeptical and by age 7 almost all children chose the box opposite to the one indicated by E1.

In another study, 5-year-olds played a similar game with a puppet in which one player's task was to find a hidden reward while the other player could provide hints about its location (Reyes-Jaquez & Echols, 2015). In a first phase, children observed an experimenter hiding a reward under one of two boxes and children were asked to tell a naïve puppet where to look for it. In one condition, children won the reward if the puppet picked the wrong box so that children had an incentive to deceive the puppet. In another condition, players' interests were congruent so that both won a reward if the puppet searched in the correct location. In a second phase, the roles were switched: The puppet now observed the hiding event and provided a hint while children assumed the role of the guesser. Children mistrusted the puppet more if their interests were conflicting than when their interests were congruent suggesting that children took into account the situational incentives when deciding whether or not to trust her advice.

While these results are intriguing, children in the incongruent condition may have learned in the first part of the study that picking the opposite box as indicated by the informer is a successful strategy for the guesser. Indeed, participants' dishonesty in the first part and mistrust in the second part of the study were correlated (playing the role of the informant first also modified children's trust decisions in the study by Sher et al., 2014). Moreover, the informant's hints in these two studies were not provided spontaneously. Instead, they were part of the game instructions and participants were made fully aware that it was up to the informer to offer either correct or false information. This may have sensitized children to the possibility of being deceived. Hence, while these studies demonstrate that already late preschool children are able to consider other's incentives in some contexts it is not clear if they can do this spontaneously and in the absence of prior cues.

In the current study, we therefore aimed at presenting children with a scenario reflective of communication in everyday life in the sense that children had to decide whether to trust utterances which were provided entirely spontaneously. Moreover, children were not informed about the intentions of the informant and instead had to independently infer them from incentives alone. We were interested in whether and from what age on children mistrust hints by individuals whose incentives are directly opposed to their own compared to individuals whose incentives are congruent. In addition, we explored how children adjust their trust decisions in response to

receiving misleading advice and if their interpretations of this advice depend on the informant's motives. For instance, wrong hints by agents with opposing incentives might be interpreted as intentionally deceptive leading children to mistrust that agent in the future, whereas the same advice by an agent with congruent incentives may be interpreted as an unintentional mistake. Alternatively, children may attach greater importance to outcomes of others' prior testimony than to their incentives resulting in mistrust towards an agent who had misled them before regardless of the agent's incentives (Liu et al., 2013).

To address these issues, we presented young children with a ball-finding task. Participants were asked to search for a ball which was hidden in one of two boxes. A co-player who had visual access to the hiding event had either congruent or opposing incentives as the child. That is, the co-player benefitted either from the child's success or failure in the ball-finding task. At test, children were given a spontaneous and unsolicited hint about the location of the ball by a co-player. We tested whether children, in the complete absence of prior hints, differentially mistrusted the co-player in a cooperative (i.e. same interests) and a competitive (i.e. opposite interests) game context. We also examined whether, after being misled once, children's decisions to trust changed on a subsequent trial and whether this differed by condition. To further investigate children's interpretations of the misleading advice we finally asked children to verbally evaluate the test situation.

In the cross-sectional design, we tested children at age 5 and 7. These age groups were chosen because previous research has indicated first competencies at around age 5 in situations in which children received additional scaffolding to assess the informant's trustworthiness from situational incentives (e.g. prior experience in taking the informant's role, direct information about the informant's malevolent intent, or incorporation of the hint as a part of a game procedure). Since we expected it to be more difficult for children to evaluate spontaneous hints we decided to test both 5-year-olds as well as an older age group for comparison.

2. *Materials and Methods*

2.1 Participants

A total of 47 five-year olds ($M_{\text{age}} = 5$ years 7 months; range = 5 years 4 months 3 days – 5 years 9 months 26 days) and 48 seven-year-olds ($M_{\text{age}} = 7$ years, 6 months; range = 7 years, 4 months, 4 days – 7 years 9 months 28 days) were tested in their urban day-care centers or elementary schools. Parental consent and children’s assent was obtained prior to testing. Participants came mostly from middleclass backgrounds. Six additional participants were excluded due to experimental error and two children were excluded as they failed to pass the training criteria (see below). Another 6 children started the game but did not want to continue playing during the training phase. One additional child was excluded from the analysis as it was later found out that she was four and not five years old at the date of testing.

2.2 Apparatus and Design

Two opaque boxes (12cm x 12cm x 12cm), one blue and one yellow, were placed on the right and the left hand side on top of a small table (82cm x 22cm x 53cm) underneath a Perspex case (70cm x 16cm x 16cm). A handle protruding laterally through a slit in the Perspex case was attached to each box. Children could move the boxes sideways on the table by using these handles. In the middle of the table were two holes. If one of the boxes was moved to the middle of the table any ball hidden inside it fell through a Perspex tube into a reward cup placed underneath. A red “magic curtain” (130 cm x 153cm) was used to block children’s visual access to the boxes when the balls were hidden (see Fig 1 for the general set-up).

Children played a ball-finding-task with a co-player. In a between-subjects design children were randomly allocated to one of two conditions. Children’s goal in both conditions was to retrieve a ball hidden in one of the two boxes. The critical difference between conditions was the co-player’s role: In the cooperative condition, children and their co-player’s interests were identical: Both received a sticker if the child successfully retrieved a ball from one of the boxes. If children chose the incorrect box both players received nothing and the ball was instead stolen by the “trash monster” – a cardboard box placed underneath the tube into which the ball fell once the correct box was moved into the middle of the table. In the competitive condition, children and their co-player had opposing interests: If children chose the correct box, they received a sticker while their co-player received nothing. If they chose the incorrect box, on the other hand, their co-player received a sticker and the child received nothing.

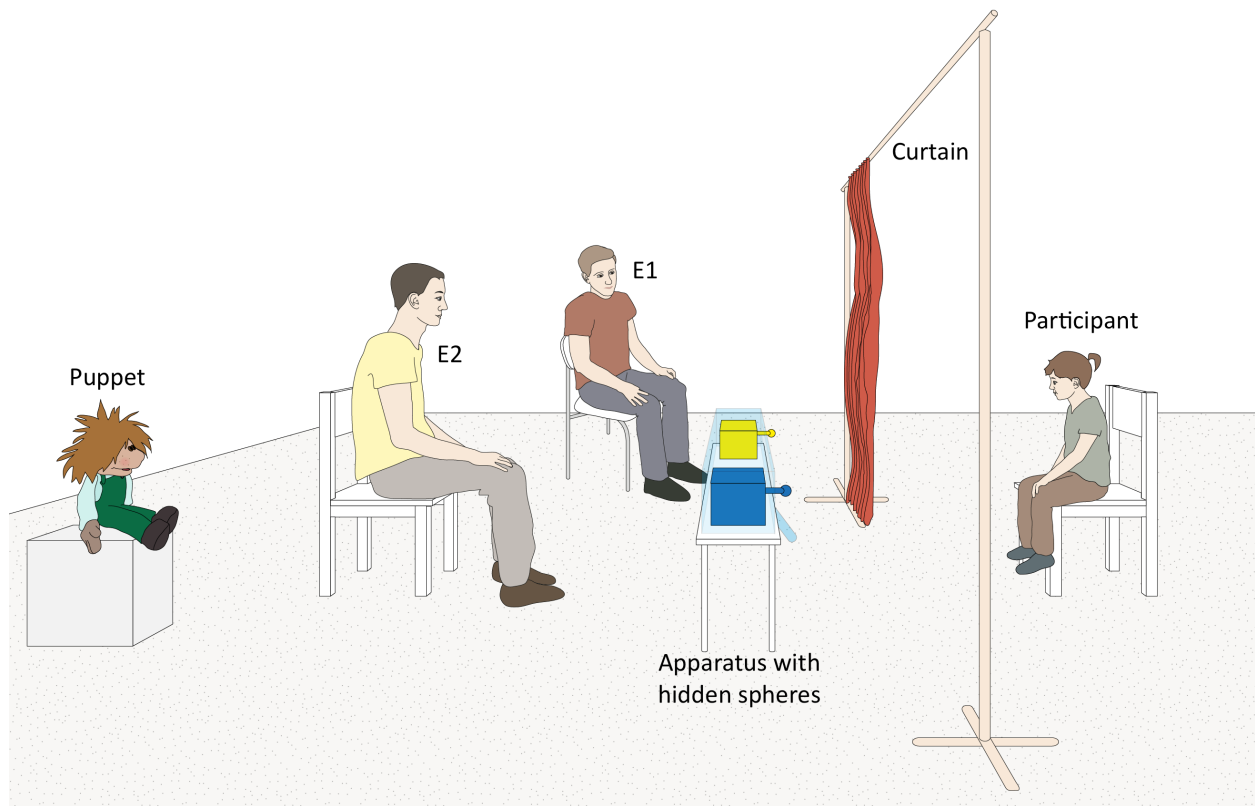


Figure 1. General set-up of the study

2.3 Procedure

Children first played six training rounds to become familiarized with the setup and to get used to the ball-finding game including their own and the co-player's incentives. During these trials, children had to guess the location of the ball. Before test, children were asked comprehension questions to guarantee that they understood the game. At test, children engaged with a new co-player who provided children with a spontaneous hint about the location of the ball (first test trial). In all cases, this hint was misleading, so that children only succeeded if they mistrusted the co-player. In a second test trial, children were again given an unsolicited hint by the co-player. After the two test trials, children were asked about their own evaluation and interpretation of the misleading hints of the co-player.

2.3.1 Training

The study started with some training trials to ensure that participants fully understood the game's incentives. A first experimenter (E1) introduced the children to a puppet and the second experimenter (E2). At training children played with E2 while the puppet was moved to the back of the room. E1 told the players that they could win stickers if they successfully retrieved a ball hidden in one of the two boxes. In the first round, there were no top covers on the boxes, so that both players could see where the ball was. Children then moved the correct box towards the middle of the table so that the ball fell into the child's reward cup. Children then exchanged the ball for a sticker from E1. Depending on the condition E2 was also rewarded (cooperative condition) or not (competitive condition) and uttered excitement or disappointment, respectively. Before re-baiting the apparatus, E1 closed the curtain between children and the boxes, so that participants could not see E1 hiding the balls. Moreover, E1 pointed out that E2 would see the hiding event and would therefore know the true location of the ball. E1 then placed a ball in both boxes but rigged the apparatus such that no ball could fall through the tube. Hence, regardless of which box they chose children failed to retrieve the ball. E1 then opened the curtain and asked E2 to leave the room in order to allow participants to decide on their own which box to choose. Upon leaving, E2 stated "hopefully I will win a sticker". This statement served to highlight the puppet's motivation for winning stickers and did not vary between conditions. While the participants made their decisions, E1 turned away. When children failed to retrieve the ball E1 stated that now E2 (competitive condition) or "trash monster" (cooperative condition) could steal the stickers. E2 then re-entered and pointed out that the ball must have been in the opposite box, while acting happily or sadly depending on condition. E1 secretly unrigged the apparatus after which E2 or the "trash monster" retrieved the ball by moving the box not chosen by the participants to the middle of the table. Children then played four more rounds in which the apparatus was manipulated so that children won and failed to win the sticker in alternating fashion (in accordance with the 50% chance level). This was done to ensure equal training experiences across participants and to demonstrate that the location of the ball was randomized. In these rounds, children had to guess the location of the ball and did not receive any hints by either E2 or E1.

After completion of the six training trials E2 collected his rewards and left the room. Once E1 and participants were alone in the room, E1 asked children the following questions: “Did you win a sticker last round? [...] And how about E2, does he also get a reward if you find the ball? [...] But what happens if you don’t find the ball? [...]” If participants failed to answer these questions correctly, they were excluded from further analyses.

2.3.2 Test

At test, the puppet (played by E2) was introduced as the new co-player. We decided against using an adult partner to prevent children from trusting the partner’s hint merely out of deference or politeness (indeed, previous research has shown that children trust adults more than peers; see Jaswal & Neely, 2006). An initial attempt to use peer stooges was abandoned since they proved to be too unreliable when providing the hint and too inconsistent in their interactions with target children. We therefore used a puppet instead, which has been a common technique in previous studies on children’s epistemic trust (e.g. Mascaro & Sperber, 2009; Liu et al., 2013; Reyes-Jaquez & Echols, 2015).

As in training trials, the co-player but not the participants could observe E1 hiding the ball. Now, however, only one of the two boxes was baited. The location of the ball in the first test trial was counterbalanced between participants. After opening the curtain, E1 said that he/she had to leave the room briefly and asked the two players not to continue with the game until his return. After E1 left the room, the puppet provided an unsolicited misleading hint directing children to the empty box (“Psst, the ball is in the blue/yellow box.”). As in the training trials, he added “hopefully, I will win a sticker!” (see Fig 2a). This was the first hint children received in the game and, from the child’s perspective, was provided entirely spontaneously. E1 then returned and – as in all previous trials – asked the co-player to leave the room, so that participants could decide in private (see Fig 2b). Children found the ball if they mistrusted the puppet’s hint but not if they trusted the puppet. The test trial was then repeated in identical fashion except that the puppet’s hint indicated the opposite box and that this time both boxes were baited such that children received a reward regardless of their choice.

Following the second test trial, E1 asked children several questions to assess their understanding of the test situation and their interpretation of the puppet's hints. E1 first asked whether the puppet told them anything in his/her absence. E1 further asked what the puppet had told them and whether it was true or false. If children indicated that the puppet's hint was false, E1 finally asked children why they thought the puppet's hint was false. Following the questions, children were told that the puppet made a mistake and did not want to deceive them. The puppet then came in and apologized for his misleading hints and children received the stickers they won in the experiment.

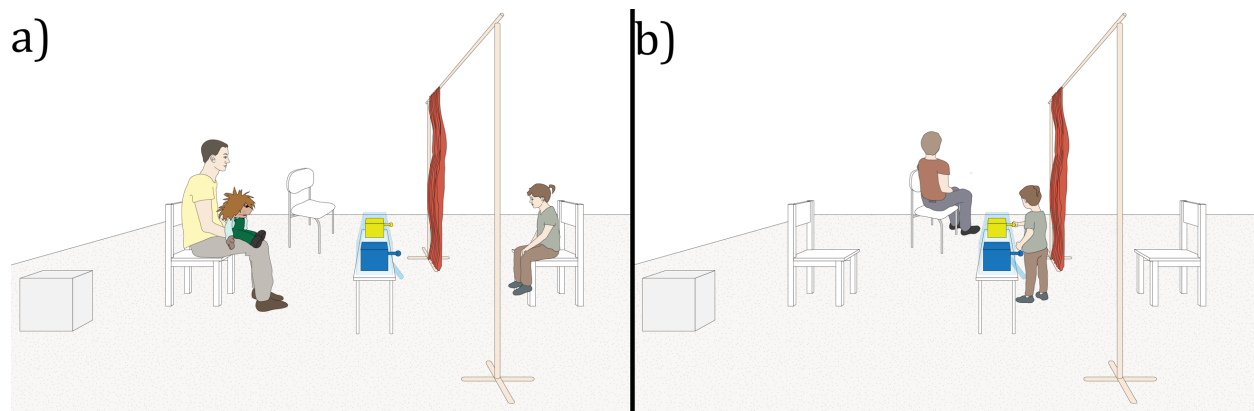


Figure 2. Critical situations of the study: a) deceptive hint by the puppet in the absence of E1; b) participant's decisions in the absence of E2 and the puppet, E1 turning away

2.4 Coding

All experimental sessions were videotaped and coded from tape. We coded children's choice of box in all training trials (i.e. blue or yellow), whether they trusted or mistrusted the puppet's hints in the two test trials, and how children interpreted the puppet's misleading hints during the post-test verbal assessment. Children's answers were coded as "deceptive" if they indicated that the puppet gave the hint with deceptive intent (e.g. "He tried to trick me.") and as "non-deceptive" if children did not indicate that they felt tricked or intentionally misled. A second coder blind to the study's predictions recoded the data. Inter-rater reliability was very

high ($\kappa = 1$ for children's choice of box, and $\kappa = 0.801$ for the verbal assessment). The data was analyzed in R (R Development Core Team, 2015).

3. Results

3.1 Spontaneous Mistrust

We first examined how many children mistrusted the puppet’s hint in the competitive condition as compared to the cooperative condition on the first test trial. Children of both age groups mistrusted a puppet with conflicting incentives more than a puppet with congruent incentives (5-year-olds: Fisher’s exact test, $p = .009$, 7-year-olds: Fisher’s exact test, $p = .002$). Children’s mistrust in both round 1 and 2 is summarized in table 1.

Table 1

Mistrust in the cooperative and competitive condition

		Test Trial	Condition	
			Cooperation	Competition
Age	5-year-olds	1	0 (24)	6 (23)
		2	13 (24)	20 (23)
	7-year-olds	1	1 (24)	11 (24)
		2	21 (24)	22 (24)

Note. In brackets: total number of children tested in the respective condition and age group; test trial 1 is indicative of children’s spontaneous mistrust; test trial 2 shows children’s mistrust after receiving a misleading hint after round 1

3.2 Children’s overall trust decisions

To examine whether children’s decisions to trust changed differentially over trials depending on the condition and to test for overall condition and age effects a Generalized Linear Mixed Model (GLMM; Baayen, 2008) with binomial error structure and log link function (McCullagh & Nelder, 1989) was fitted, using the function *glmer* from the package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). Condition, trial number, age, and their two- and three-way interactions were included as fixed effects “test predictors” (Mundry, 2014). As the informant’s

sex has previously been argued to influence children's trust behavior in that children show a same-gender bias (Ma & Woolley, 2013), sex was included as an additional fixed effect. Finally, subject identity was included as a random effect. The response measure was whether participants trusted or mistrusted the puppet. Due to complete separation issues (Field, 2005) – all 5-year-olds in trial 1 of the cooperative condition trusted the puppet – the estimates for this model could not be calculated reliably for the original data. Therefore, each individual 1 in trial 1 of the cooperative condition was replaced by a 0, one at a time (with 1 representing “trust” and 0 “mistrust”) and for each replacement a separate model was fitted to the derived response (for estimating the null model, five of these models had to be excluded as extreme variances led to unreliable models). For the overall result, all converging models were averaged (due to inflated standard errors in some of the models median rather than mean estimates are reported). It is worth noting that this approach leads to somewhat more conservative estimates and p-values. Comparing the full model with a null model not including the test predictors revealed a significant effect ($\chi^2 = 104.16$, $df = 7$, $p < .001$) suggesting that the test predictors combined affected trusting decisions. The full model was then compared to a reduced model not including the three-way interaction. The interaction was not significant (median estimate \pm SE = -0.70 ± 0.60 , $\chi^2 = 1.288$, $df = 1$, $p = .261$) and was consequently dropped from the model. Further likelihood ratio tests revealed no significant interactions between condition and age (-0.208 ± 0.589 , $\chi^2 = 0.126$, $df = 1$, $p = .722$), condition and trial number (-0.889 ± 0.585 , $\chi^2 = 2.67$, $df = 1$, $p = .102$), and trial number and age (-0.208 ± 0.589 , $\chi^2 = 0.238$, $df = 1$, $p = .626$). However, analyses of the main effects revealed that, overall, children mistrusted the informant more in the competitive condition than in the cooperative condition (1.985 ± 0.623 , $\chi^2 = 17.214$, $df = 1$, $p < .001$), 7-year-olds mistrusted more than 5-year-olds (-0.656 ± 0.281 , $\chi^2 = 83.02$, $df = 1$, $p < .001$), and children mistrusted more on the second than on the first trial (-1.984 ± 0.510 , $\chi^2 = 7.774$, $df = 1$, $p = .005$; see Figure 3). There was no significant sex difference in children's trust decisions (0.323 ± 0.476 , $\chi^2 = 0.615$, $df = 1$, $p = .454$).

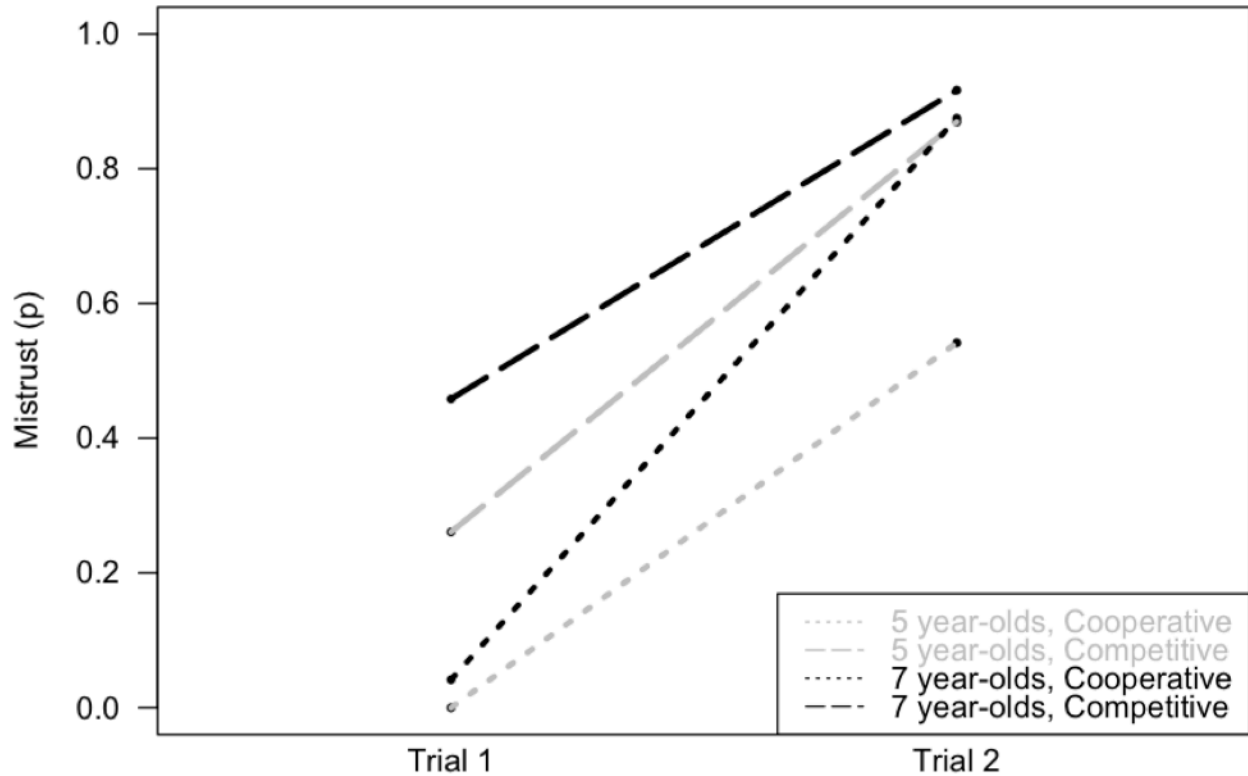


Figure 3. Children's mistrust in hints given by the puppet

3.3 Interpretation of Misleading Information

Both the 5-year-olds (Fisher's exact test, $p = .003$) as well as the 7-year-olds (Fisher's exact test, $p < .001$) were more likely to attribute deceptive intentions to the puppet in the competitive than in the cooperative condition. To follow up this result we further ran a generalized linear model (GLM) with binomial error structure¹. Age group, condition, and their interaction were included as predictors. Interpretation of the misleading hints (deceptive vs. non-deceptive) was included as a dependent variable. This model was then compared to a reduced model not containing the interaction between the two predictors. The interaction was not significant (estimate \pm SE = -0.54 ± 0.65 , $\chi^2 = 0.755$, $df = 1$, $p = .385$) and was thus dropped from the model. The significance of each main effect was then estimated comparing the model with a model not containing the

¹ We thank an anonymous reviewer for suggesting this analysis. Since we did not have clear predictions regarding the interaction and the effect of age this analysis should be regarded as explorative.

respective predictor. This revealed that older children and children in the competitive condition were more likely to interpret the puppet’s hint as intentionally misleading than younger children and children in the cooperative condition, respectively (effect of age: 0.94 ± 0.28 , $\chi^2 = 13.390$, $df = 1$, $p < .001$; effect of condition: -2.42 ± 0.57 , $\chi^2 = 23.512$, $df = 1$, $p < .001$). The data is shown in Table 2.

Table 2

Interpretation of misleading information

		Condition	
		Cooperation	Competition
Age	5-year-olds	4 (24)	14 (23)
	7-year-olds	11 (23)	21 (22)

Note. Number of children interpreting the informants misleading hint as intentionally deceptive; In brackets: total number of children tested in the respective condition and age group

4. Discussion

Both 5-year-olds and 7-year-olds mistrusted a hint given by a competitive co-player more often than a hint given by a cooperative co-player. This confirms previous findings suggesting that it is within the capacity of five-year-old children to take into account others’ incentives when evaluating the truthfulness of their statements (Sher et al., 2014; Reyes-Jaquez & Echols, 2015). In contrast to previous investigations, however, children in the current study had to infer the informant’s intention to deceive entirely from his/her incentives as afforded by the game situation and they received the informant’s hint spontaneously and without any pre-warning. Moreover, children in the competitive condition were more likely to interpret the misleading hints of a co-player as deceptive than children in the cooperative condition. This underlines that children did indeed consider their partner’s incentives when assessing the reason for the co-players misleading hint.

However, while this indicates first competencies at evaluating utterances based on the informant's incentives already at five years of age the majority of children trusted the informant even in the competitive condition. Indeed, only about half of the 7-year-olds spontaneously mistrusted the co-player whose interests were at conflict with their own indicating that children may at first assume to be provided with truthful information even in competitive contexts. This finding appears to contrast with previous studies. Already around age 4, children have been shown to mistrust informants based on their (malevolent) dispositions (Mascaro & Sperber, 2009; Heyman et al., 2014; Landrum et al., 2013) and by age 5, they consider whether an informant has previously shown deceptive intent (Liu, 2013; Vanderbilt, 2011; Mascaro & Sperber, 2009). Our results thus highlight the importance of a clear distinction between informants' incentives and their intentions and motives for the investigation of deception and selective trust. Although preschoolers are generally capable of integrating intent-based information into their trust decisions (see also Shafto et al., 2012), our study indicates that the ability to infer trustworthiness from situational incentives alone emerges somewhat later in development.

Compared to previous studies (Sher et al., 2014; Reyes-Jaquez & Echols, 2015), these results suggest a more conservative appraisal of preschoolers' abilities for incentive-based skepticism. This discrepancy may be due to the fact that in these previous studies the rules of the game specified that misleading hints were possible which may have sensitized children to the risk of deception. Moreover, in the study by Reyes-Jaquez and Echols (2015) participants played the game in the role of the observer first and thus had the opportunity to learn that picking the opposite box as indicated by the informer would be a successful strategy. Hence, while the current findings confirm that it is within the capacity of preschool children to detect and mistrust informants with competing incentives, this ability appears to be less robust, especially when they have to do so independently and in the absence of any prior experience with the test situation. The default assumption of preschoolers regarding trustworthiness of informants seems to be primarily cooperative (see also Jaswal et al., 2010).

A further aim of the current study was to investigate whether children respond differently to being misled by a cooperative versus a competitive co-player. The results showed that both 5- and 7-year-old participants largely mistrusted their co-player after they found out that the first hint was false. This was the case irrespective of condition, indicating that past accuracy may

have a more salient and stronger influence on children's decisions to trust than the informant's incentives. In test trial 1, in which children had not previously been exposed to hints by the co-player, children had to base their trust decisions solely on the co-player's incentives, and thus, her assumed intentions. In test trial 2, by contrast, children appeared to primarily base their trust decisions on the co-player's past accuracy while disregarding her incentives. This corresponds to other studies suggesting that children prioritize information on past behavior over intentions when deciding whom to trust (Liu et al., 2013) and which actions to imitate (Corriveau, Min, Chin, & Doan, 2016).

Another potential reason for the similar increases in children's mistrust in both conditions irrespective of the co-player's incentives is that the trust measure of the current study may not have been sufficiently sensitive. Measuring trust using a graded scale rather than a categorical decision could potentially test whether children are able to integrate their partner's past accuracy with information about his or her incentives into their trusting decisions. Indeed, the ways in which children weigh different sources of information throughout ontogeny would be an interesting topic for future investigations. It would also be interesting to examine children's trust in a co-player whose self-interest changes during the game to find out how flexibly they can adjust their trusting behavior in response to their interlocutor's incentives

When interpreting the co-player's hints, both 5- and 7-year-olds were more likely to attribute dishonest intentions when the co-player's incentives were competitive than when they were cooperative. Of those children who did not explicitly interpret the hint as dishonest in the cooperative condition, most children said they did not know why the puppet gave an incorrect hint (note that there was indeed no obvious reason for the puppet to lie in the cooperative condition) while some children, especially 5-year-olds, came up with different explanations (e.g., that the puppet did not know the correct location or was just silly, or they wrongly insisted that the hint had been correct). Neither the 7- nor the 5-year-olds interpreted the puppet's misleading hint in the cooperative condition as an unintentional mistake. This may explain their reluctance to take into account the puppet's incentives and their tendency to instead focus on past accuracy on trial 2. It would be interesting to see if children consider an informant's incentives more in situations in which the misleading advice can easily be excused by mistakes (e.g. when it was difficult for the informant to get the required information or the hint was was not based on first-hand information, see also Nurmsoo & Robinson, 2009a; 2009b).

Finally, our data suggests that children's sensitivity to the risk of deception continues to develop with age. One ability that is related to selective trust in young children is mind-reading (Vanderbilt et al., 2011; Brosseau-Liard, Penney, & Poulin-Dubois, 2015; DiYanni, Nini, Rheel, & Livello, 2012). Since deception involves the purposeful implantation of a false belief into others (A wants B to have a false belief about the location of the ball), higher-order mind reading abilities may be crucial for such selective mistrust (e.g. B knows that A wants her to falsely believe that the ball is in the blue box). This is also one plausible explanation for the 5-year-olds' difficulties when handling spontaneously elicited hints by a competitor. As children become better at reasoning recursively about mental states they may thus become better at spontaneously detecting when others intend to install a false belief in themselves. A potential avenue for future research could therefore be to investigate higher-order mind reasoning skills and its relation to selective trust. Interestingly, one 7-year-old girl in the current study, after mistrusting the first hint, followed the second hint of the puppet because she expected the puppet to anticipate her mistrust and to therefore provide a truthful hint on the second trial (as indicated by her verbal justification). This corresponds to previous findings showing that children at this age increasingly apply sophisticated mind-reading capabilities in social interaction (Grueneisen, Wyman, & Tomasello, 2015; Filippova & Astington, 2008) and this may also facilitate their ability to successfully evaluate others' credibility (see also Sher et al., 2014).

Behavioral inhibition might also be a crucial prerequisite for detecting deceptive advice. Children may have to learn to inhibit the impulse to unquestioningly follow other's hints and to integrate several relevant pieces of information (e.g. other's incentives, their past behavior, etc.) into their trust decisions. Examining these cognitive capacities in relation to children's selective trust may thus help to further elucidate the mechanisms underlying the development of this facility. A third factor contributing to children's developing mistrust may also be social experience. During the preschool and early school years, children are likely to be increasingly exposed to peer interactions involving deception. Since children are also becoming better deceivers with age (e.g. Evans, Xu, & Lee, 2011), the risk of being deceived may also increase during that time. Developing selective mistrust may thus go hand in hand with increasing peer interactions during the preschool years. As pointed out by the Reviewer, integrating such factors in future research will help us to further understand the ontogeny of selective mistrust.

One potential limitation of the current study is that the child had to decide whether to trust information given by an adult experimenter. Mistrusting an authority figure is likely to be more difficult for children than trusting, for example, peers. Ideally, peer confederates could work as informants in future studies. We initially piloted the paradigm using peer stooges but refrained from this method in order to maximize the standardization of the procedure. We used a puppet instead to decrease the likelihood that children trusted merely out of deference. As indicated by children's high levels of mistrust in test trial 2, children did indeed seem comfortable to mistrust the puppet. One related concern is that the puppet was played by E2 with whom children already engaged in training trials. In this phase, however, the co-player did not give any hints about the ball's location. Thus, prior experience with E2 is unlikely to systematically account for children's trust decisions in the game. Nonetheless, introducing novel, entirely neutral characters as informants would have been an ideal solution to rule out any effects of prior experience.

In summary, the current study found that from five years of age children are capable of interpreting misleading hints depending on the incentives of the informant which demonstrates that by the late preschool years children possess the cognitive capacity to be vigilant when faced with the risk of deception. However, this ability is not uniformly displayed: most preschoolers are trustful even when dealing with advice provided by informants with conflicting incentives. Indeed, only about half of the 7-year-olds mistrusted a competitive partner in the absence of other information about the informant. Compared to others' incentives, prior personal experience seems to play a crucial role in children's mistrust: after experiencing signs of deception, children immediately switch to a mistrusting attitude regardless of the incentive structure of their interlocutors. Children's skills at detecting deceivers based on their incentives are thus still developing well into middle childhood.

5. References

- Baayen, R. H. (2008). *Practical data analysis for the language sciences with R*. Cambridge, MA: Cambridge University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Brosseau-Liard, P., Penney, D., & Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *British Journal of Developmental Psychology*, 33(4), 464–475. <https://doi.org/10.1111/bjdp.12107>
- Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going With the Flow: Preschoolers Prefer Nondissenters as Informants. *Psychological Science*, 20(3), 372-377. doi: 10.1111/j.1467-9280.2009.02291.x
- Corriveau, K. H., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, 12(1), 188-193. doi: 10.1111/j.1467-7687.2008.00763.x
- Corriveau, K. H., Min, G., Chin, J., Doan, S. (2016). Do as I do, not as I say: Actions speak louder than words in preschoolers learning from others. *Journal of Experimental Child Psychology*, 143, 179-187. doi: 10.1016/j.jecp.2015.10.006
- DiYanni, C., Nini, D., Rheel, W., & Livelli, A. (2012). “I Won’t Trust You if I Think You’re Trying to Deceive Me”: Relations Between Selective Trust, Theory of Mind, and Imitation in Early Childhood. *Journal of Cognition and Development*, 13(3), 354–371.
- Evans, A.D., Xu, F., & Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology*, 47(1), 39-49. doi: 10.1037/a0020787
- Field A. (2005). *Discovering Statistics using SPSS*. London: Sage Publications.
- Filippova, E. & Astington, J. (2008). Further Development in Social Reasoning Revealed in Discourse Irony Understanding. *Child Development*, 79(1), 126-138. <http://dx.doi.org/10.1111/j.1467-8624.2007.01115.x>
- Fusaro, M. & Harris, P. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, 11(5), 771-777. doi: 10.1111/j.1467-7687.2008.00728.x

- Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know..." Children use second-order false-belief reasoning for peer coordination. *Child Development, 86*, 287-93. doi: 10.1111/cdev.12264
- Henrich, J. (2015). *The Secret of Our Success: How culture is driving human evolution, domesticating our species, and making us smart*. Princeton, NJ: Princeton University Press.
- Heyman, G. D., Sritanyaratana, L., & Vanderbilt, K. E. (2013). Young Children's Trust in Overtly Misleading Advice. *Cognitive Science, 37*(4), 646-667. doi: 10.1111/cogs.12020
- Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*(10), 1541- 1547. doi: 10.1177/0956797610383438
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: preschoolers use past reliability over age when learning new words. *Psychological Science, 17*(9), 757-758. doi: 10.1111/j.1467-9280.2006.01778.x
- Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science, 16*(4), 622–638. <https://doi.org/10.1111/desc.12059>
- Liu, D., Vanderbilt, K. E., & Heyman, G. D. (2013). Selective trust: Children's use of intention and outcome of past testimony. *Developmental Psychology, 49*(3), 439-445. doi: 10.1037/a0031615
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*(3), 367-380. doi: 10.1016/j.cognition.2009.05.012
- Ma, L. & Woolley, J. (2013). Young Children's Sensitivity to Speaker Gender when Learning From Others. *Journal Of Cognition and Development, 14*(1), 100-119. doi: 10.1080/15248372.2011.638687
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models* (2 ed.). New York: Chapman & Hall.
- Mundry, R. (2014). Statistical Issues and Assumptions of Phylogenetic Generalised Least Squares (PGLS). In Garamszegi, L. (Ed.). *Modern Phylogenetic Methods and their Application in Evolutionary Biology*. Heidelberg: Springer.

- Nurmsoo, E., & Robinson, E. J. (2009a). Children's trust in previously inaccurate informants who were well or poorly informed: when past errors can be excused. *Child Development*, 80(1), 23-27. doi: 10.1111/j.1467-8624.2008.01243.x
- Nurmsoo, E., & Robinson, E. J. (2009b). Identifying unreliable informants: do children excuse past inaccuracy? *Developmental Science*, 12(1), 41-47. doi: 10.1111/j.1467-7687.2008.00750.x
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, 43(5), 1216-1226. doi: 10.1037/0012-1649.43.5.1216
- R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Reyes-Jaquez, B., & Echols, C. H. (2015). Playing by the rules: Self-interest information influences children's trust and trustworthiness in the absence of feedback. *Cognition*, 134(0), 140-154. doi: 10.1016/j.cognition.2014.10.002
- Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: modeling children's reasoning about others' knowledge and intent: Modeling epistemic trust. *Developmental Science*, 15(3), 436-447. <https://doi.org/10.1111/j.1467-7687.2012.01135.x>
- Sher, I., Koenig, M., & Rustichini, A. (2014). Children's strategic theory of mind. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37), 13307-13312. doi: 10.1073/pnas.1403283111
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, 25(4), 359-393. doi: 10.1111/j.1468-0017.2010.01394.x
- Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The Development of Distrust. *Child Development*, 82(5), 1372-1380. doi: 10.1111/j.1467-8624.2011.01629.x