



The development of intent-based trust in moral testimony

Francesco Margoni^{1,2} · Elena Nava³

Accepted: 23 April 2024
© The Author(s) 2024

Abstract

Research has shown that by age 5–6 years, children fully integrate information about agents' mental states into their verbal moral judgments: When asked to say whether an agent is morally good or bad, they rely on the agent's intentions more than on its action's outcomes. Research has also shown that from an early age, children use a plethora of social and moral cues when deciding whom to trust in learning and testimony situations. Here, for the first time, we asked if and how children's trust in informants who relay information about the moral character of a novel agent is influenced by the valence of the intentions underlying the informants' prior actions. Italian children aged 6 to 10 years ($n=219$, 112 female) were first presented with two puppets and asked to judge them. One puppet accidentally caused harm (neutral intention, negative outcome), the other attempted but failed to do so (negative intention, neutral outcome). Next, the puppets gave contrasting testimonies about whether a novel agent was good or bad. Findings revealed that the tendency to trust the assessment of the well-intentioned puppet concerning the novel agent emerged at age 8, whereas younger children simply showed to believe that the novel agent was good, regardless of the testimonies they received. These results suggest that despite the ability to generate intent-based moral judgments emerges at age 5–6, the tendency to rely on intentions underlying past actions of informants when assessing informants' testimonies about the moral character of a third party undergoes significant change in childhood.

Keywords Epistemic trust · Testimony · Moral judgment · Trait attribution · Childhood

Human societies depend on cultural and knowledge transmission and cooperative exchanges (Henrich, 2016). Individuals, and especially children, rely on others' testimonies to successfully navigate their social and epistemic environment. Because of this dependency on others in acquiring knowledge, a serious pitfall is the possibility of being deceived. Thus, individuals tend to exercise *epistemic vigilance*, that is, they beware of potential sources of misinformation by assessing what are the intentions and aims of informants and, ultimately, whether they are likely to be honest (Mercier, 2020; Sperber et al., 2010). Here, for the first time, we asked if school-age children (6–10

years) exercised epistemic vigilance in a task that required them to determine the moral character of a novel agent. Children were presented with two informants who first demonstrated to possess either positive or negative intentions toward a third party and subsequently provided contrasting testimonies about the moral character of a novel agent; children were asked to decide which moral testimony to trust.

Epistemic trust

Deciding whom to trust is a crucial task, and children's *epistemic trust* has been the subject of a number of studies. Whereas infant researchers have provided evidence that preverbal infants hold physical, numerical, biological, linguistic, psychological, and sociomoral expectations (sometimes defined naïve theories or core knowledge) overall suggesting that humans are equipped with early-emerging and dedicated learning mechanisms for representing notions across a variety of domains (Carey, 2009; Spelke, 2022; see also

✉ Francesco Margoni
francesco.margoni@uis.no

¹ Department of Social Studies, University of Stavanger, Stavanger, Norway

² Department of Psychology, University of Oslo, Oslo, Norway

³ Department of Psychology, University of Milano-Bicocca, Milan, Italy

Margoni et al., 2024), cognitive developmental psychologists have started addressing the further question of how most of our beliefs about the world (which do not fall into the category of basic naïve built-in knowledge) are acquired via *testimony*. For example, to learn about historical events, children necessarily need to rely on the testimony of others. Developmental psychologists use the terms selective trust, trust in testimony and/or epistemic trust to denote the process of selecting whom to believe and whom to approach to gain new information, a process that allows cultural and knowledge transmission in the human species (e.g., Clément et al., 2004; Harris & Corriveau, 2011; Koenig & Harris, 2007). A crucial prediction of evolutionary models of cultural transmission (Richerson & Boyd, 2005) is that humans are equipped with early-emerging biases that guide their selection of the models or informants to learn from. Within this theoretical framework, researchers have investigated if children trust some informants systematically more than others, finding a positive answer to the question and further revealing under which conditions an informant is trusted.

Exhaustively reviewing this literature would be beyond the scope of this short article. Still, a few relevant findings can be quickly summoned for the reader. Overall, we know that from an early age, children use a plethora of cues when deciding whom to trust (for a meta-analysis, see Tong et al., 2020). When the task requires to learn a new word, an object function, or acquire a new belief, preschoolers selectively trust informants who previously had been accurate vs. inaccurate (Birch et al., 2008; Clément et al., 2004; Corriveau et al., 2013; Koenig et al., 2004; Pasquini et al., 2007; Ronfard & Lane, 2018), knowledgeable vs. ignorant (Gillis & Nilsen, 2013; Koenig & Harris, 2005; Sabbagh & Baldwin, 2001), and reliable vs. unreliable (Rakoczy et al., 2009; Ronfard & Lane, 2019). Moreover, pointing to a possible cautionary and adaptive strategy, preschoolers have been shown to attend to familiar, similar, ingroup or native-accented informants over unfamiliar, dissimilar, outgroup or foreign-accented informants (Corriveau & Harris, 2009; Elashi & Mills, 2014; MacDonald et al., 2013; Kinzler et al., 2011).

Research has also shown that children's socio-moral evaluations of informants can have downstream effects on their decisions about whom to trust and learn from (Koenig et al., 2019; Marble & Boseovski, 2020). Some studies have reported that preschoolers selectively rely on the testimony of kind (over mean) agents and of agents who kept their promises (over those who did not) (Isella et al., 2018; Mascaro & Sperber, 2009), whereas other studies have found evidence of a gradual development of these tendencies during middle childhood (Heyman et al., 2013; Vanderbilt et al., 2011). Doebel and Koenig (2013) have

also reported that preschoolers selectively trust informants who behaved in a good vs. bad manner, but effect sizes were small. And in an interpersonal trust task, school-age children but not preschoolers were shown to trust helping agents more than hindering agents (Margoni et al., 2022). However, even more than external outcomes of one's actions (i.e., whether one helped or told the truth vs. harmed or lied), the *intentions* underlying one's actions are ostensibly taken by people as particularly diagnostic of an agent's moral character (Cushman, 2015; Monroe & Malle, 2017), where the assessment of the informant's moral character has been shown to guide children's trust. Indeed, Liu et al. (2013) have demonstrated that 5- and 6-year-olds rely on their assessment of informants' past outcomes but also intentions (to help or deceive others by relaying true or false information) to decide whether to trust their testimony about objects location.

Building on this prior research, the main aim of the current study was to test— for the first time— if children also use their intent-based evaluation of the moral character of an informant to guide their decision whether to trust its moral evaluation of a novel unknown agent. We now briefly review the literature on the ontogeny of intent-based moral judgment, which will also prove useful in justifying why we focused on middle childhood in our study.

Intent-based moral judgment and mentalistic reasoning

Integrating the evaluation of people's intentions into the moral judgment of those people's actions or character traits is not an easy task for children, especially when the valence of the agent's intentions does not match the valence of the action's outcomes, as in the case of accidental harm (Margoni & Surian, 2016, 2020; Rosset & Rottman, 2014). Whereas young preschoolers, at about age 3 or 4, rely mainly on actions' outcomes or equally on intentions and outcomes, by age 5 or 6, children start to rely mainly (and systematically across situations and tasks that require a verbal response) on agents' *intentions* (e.g., Cushman et al., 2013; Li & Tomasello, 2018; Margoni & Surian, 2017; Nobes et al., 2016, 2017; Proft & Rakoczy, 2018). Importantly for the present study, this research suggests that to test if children's intent-based moral judgment (i.e., the tendency to prioritize intentions over outcomes in the moral evaluation) guides children's selective trust, we must focus on children older than 5. Moreover, the focus on middle childhood is justified by the presence of a well-known cognitive phenomenon for which children, especially until age 7–8, require less evidence before attributing positive vs. negative traits to people (a point we shall return to in the discussion).

At 6 years of age, children possess all the mentalistic reasoning (and executive functioning) abilities required to generate an intent-based moral judgment (Buon et al., 2016). First, they are able to attribute intentions to agents (an ability which already emerges in infancy; Baillargeon et al., 2016) and to build and maintain a representation of the relevant mental states of the agents (e.g., an intention to harm) during a moral judgment task. Second, they are able to integrate such representation in their verbal judgments of the agents' goodness or badness, and to suppress a prepotent response based on the negative outcomes (when evaluating the case of accidental harm) to favor an intent-based response (indeed, even young children with autism can do this if the task is simplified, cf. Margoni et al., 2019). Moreover, at 6, children also possess the mentalistic reasoning abilities necessary to attribute to others an intention to lie and/or deceive. The literature we reviewed above, in the section about epistemic trust, reveals that preschoolers are able to preferentially attend to competent, reliable, accurate and knowledgeable agents, even if it remains unclear whether children in most of these studies were simply discarding the information provided by the incompetent, unreliable, inaccurate and ignorant agents or they were judging the information provided by these agents to be false or (intentionally) deceitful. However, other studies have shown that by age 6, in tasks requiring to distinguish a lie from a joke, children express the capacity to understand deceptive intentions (Sullivan et al., 1995; Winner & Leekam, 1991; see also Cheung et al., 2015). Thus, in the context of the present study, children are old enough to engage in this complex form of mindreading (i.e., attribute to the informant the intention to generate a false belief in the receiver) and they might (or might not) use it to motivate the choice to discard the testimony given by the ill-intentioned informant.

The present study

We asked to what extent children, once they start expressing intent-based moral judgments, use their intent-based judgment of the informant's character to guide their decision whether to trust its testimony. We call this phenomenon 'intent-based epistemic trust'. The current study builds on prior research which has shown that children aged 6 years or younger generate intent-based moral judgments across a wide range of different tasks (Hilton & Kuhlmeier, 2019; Margoni & Surian, 2016), selectively trust well-intentioned informants when learning through testimony about objects location (Liu et al., 2013), and can be influenced by adults in forming new 'second-hand' moral beliefs (Li et al., 2019). We built on this research by investigating possible

developmental effects in intent-based epistemic trust occurring in school-age children (6–10 years). For the first time, we assessed not only selective trust in informants who relay information about objects' properties, but also in informants who relay information about the *moral character* of novel unknown agents (moral testimony). We tested if children's ability to infer people's moral character by attending to the intentions they display has important downstream consequences for learning, guiding children's decisions about whom to trust when acquiring (moral) beliefs.

In the task we designed, children had to decide whom to trust following the presentation of two agents characterized for having been either well- or ill-intentioned toward a third party: One accidentally harmed it (neutral intention, negative outcome), whereas the other attempted but failed to harm it (negative intention, neutral outcome). These two cases (attempted harm and accidental harm) are particularly useful in assessing participants' intent- or outcome-based moral judgment because intentions and outcomes lead to conflicting responses within each scenario. Differently, successful attempts to harm (negative intention, negative outcome) might be condemned because of both the negative intentions *and* the negative outcomes, and are therefore not diagnostic of a intent-based moral judgment. Next, the two agents (the attempted and the accidental harm-doer) became informants and provided the child with contrasting testimonies about the color of a ball (blue or white) possessed by a fourth agent and the moral character of the fourth agent (good or bad). Children had to decide whom to trust. Last, we assessed children's expectations about which informant the third party would trust, to test if children were also able to attribute to others the tendency to trust well-intentioned informants.

Given the wealth of research about the capacity of children in early and middle childhood to use numerous cues in epistemic trust tasks, a first prediction was that even the youngest children in our sample would selectively attend to the well-intentioned informant when the task requires to guess the color of a ball or whom the third party would trust. However, we expected to find evidence of developmental change in children's proclivity to endorse informants' moral evaluations. Selectively trusting well-intentioned agents to judge the moral character of an unknown agent (a decision motivated by the reasoning that morally good agents are less likely to deceive than morally bad agents) *might* be complicated by the difficulty to overcome default expectations that the child *might* hold about novel agents, namely that it is reasonable to assume that people are generally good rather than bad (indeed, there is evidence of a positivity bias in children consistent with this possibility which we will review in the discussion section).

Method

Raw data and study materials (Movies S1-S4) are available on the Open Science Framework, link: <https://osf.io/vkf5q/>. The study was not pre-registered. The research project received approval from the University of Milano-Bicocca Ethics Committee (#474).

Participants

The sample size was determined by an a-priori power analysis (G*Power) for a logistic regression testing the effect of age on children's trust. To detect at least an OR = 1.5 (i.e., a 10% units increase in trust), with $\alpha = 0.05$, power = 0.80, a normal distribution for the predictor (age), a minimum total sample of 208 participants was required. We recruited a total of 219 Italian school-age children. We included only children who were Italian speakers and were not affected by sensory or cognitive impairments. One 6-year-old failed at least one comprehension check and was for this reason excluded from the analyses. Thus, the final sample consisted of 218 children (age range: 6.07–11.18 yrs, 111 female; the full sample demographics are presented in Table 1). Children were recruited in an elementary school near Milano, serving a middle-income population.

Design

Children completed a task which had two parts. First, a familiarization phase in which two puppets that were shown on a laptop monitor in short movies, displayed either positive or negative intentions, causing either negative or neutral outcomes respectively (modeled after Margoni & Surian, 2020), followed by the experimenter's request to judge these puppets as good or bad. Second, the puppets became informants and children received three epistemic trust questions aimed at testing selective trust in testimony. With this design, we aimed to test children's propensity to generate intent-based moral judgments and to use such judgments to guide their decisions about whom to trust. The task mirrors real-world situations that children might have encountered, where one first judges potential informants along some

socio-moral dimensions and next, based on this evaluation, decides if they should be trusted.

Materials and procedure

Character-familiarization phase

The aim of this phase was to characterize two puppets (future informants) as either well- or ill-intentioned toward a third party, and have the child morally judge them. First, before each character-familiarization movie, children saw a 50-s introductory movie where the puppets (lion, giraffe, and dog/cat) greeted the viewer (Movie-S1-S2). After each introductory movie, children were given a *practice question* aimed at familiarizing them with the process of receiving a question and generating a response (Fig. 1). Children had to choose in which box to insert the picture of the stuffed dog/cat they just saw in the movie, whether in the 'woof box' or in the 'meow box' (the experimenter asked, "Does this animal go woof or meow? Put it in one of these two boxes. Look! This is the box of the animals which go woof, and this is the box of the animals which go meow. Which box should we put it in?"). Children were presented with a PowerPoint slide displaying in the middle a picture of the dog/cat, and on opposing sides a picture of a box with a doghouse on it and a picture of a box with a ball of wool on it. Children were encouraged to point at the right (correct) box and, by using a touchpad cursor, the experimenter moved the dog/cat's picture into the box the child pointed at.

Next, children saw two 70-s character-familiarization movies. In the *accidental harm* one, a lion puppet tripped over a stuffed dog/cat unintentionally knocking down a tower made of blocks previously built by a giraffe puppet (Movie-S3). In the *attempted harm* movie, a second lion attempted but failed to knock down the giraffe's tower, the lion used a hammer to smash it down but the tower remained intact and the lion finally desisted (Movie-S4). To help children distinguish the two lions, one wore a hat and a tie. We used puppets and simplified scenarios as these have been shown to be valid tools to test children's cognition (e.g., Lillard, 2022; Rakoczy, 2022). Moreover, in our case, the use of puppets helped reduce the potential stress related to witnessing moral transgressions and acts of property damage

Table 1 Sample demographic characteristics by age group and proportions of children who generated intent-based judgments and trusted the well-intentioned informant by age group and type of test question

Age group	Demographics		Moral judgment questions			Epistemic trust questions		
	N (N _{Female})	M _{Age} (SD _{Age})	Acc Harm	Att Harm	Two-Puppets	Q1	Q2	Q3
6 yrs	36 (21)	6.51 (0.22)	97	97	97	92	78	58
7 yrs	33 (13)	7.53 (0.33)	97	97	100	91	76	58
8 yrs	50 (25)	8.53 (0.32)	100	100	100	92	96	76
9 yrs	52 (28)	9.50 (0.26)	100	96	98	96	92	86
10 yrs	47 (24)	10.60 (0.38)	98	94	94	96	89	83

and, importantly, was further motivated by our willingness to directly follow up a prior study on children's intent-based moral judgment by using the same stimuli (Margoni & Surian, 2020). In that study, the same animal (lion) was used for both puppets (accidental harm-doer and attempted harm-doer) to rule out the possibility that children's moral judgments could be driven by a preference for a specific animal species.

After each character-familiarization movie, children received two *comprehension probes* about whether the lion destroyed the tower (*outcome question*, "Did the lion destroy the tower?") or intended to destroy the tower (*intention question*, "Did the lion intend to destroy the tower?"). If the child failed at least one probe, she was asked both questions again (this time, while making the questions, the experimenter focused the attention of the child on the last frame of the movie); if the child still failed, she was presented with the movie a second time and with the probes one last time.

Children were then presented with a PowerPoint slide displaying at its centre the picture of the lion puppet previously shown in the movie and, on opposite sides, the box 'where the good lions belong' and the box 'where the bad lions belong' (Fig. 1; an example of these stimuli can be found on the OSF SM). Children were asked to say or point at the box in which the lion should be inserted (i.e., in the 'good box' or in the 'bad box'). The experimenter said, "Look! This is the box where the good lions belong, and this is the box where the bad lions belong."

Q0 Moral judgment question: "Is this lion good or bad? Put it in one of the two boxes."

Last, after watching both character-familiarization movies, children also received the *two-puppets question*. They were presented with the same slide used for the moral judgment question but both lions were now present. Children were asked to put the lions in the boxes ("We have the two lions here. Please put them in the boxes they belong to"). Here, children could compare the lions before deciding where to put them. This test had two main aims: first, it provided a robustness check, namely the opportunity to check the internal consistency of children's answers; second, it gave us a second measure especially useful for assessing the judgment of those children who might have provided the same answer across the two moral judgment questions (indeed, in prior research, when presented with this same two-puppets question, children always used both boxes, thus they never inserted the two lions in just one box; see Margoni & Surian, 2020).

The following factors were counterbalanced across participants: the order of character-familiarization movies

(accidental harm or attempted harm first); the lions' appearance (the well-intentioned or the ill-intentioned lion wore hat and tie); the side animal (a dog or a cat was in the familiarization movie).

Epistemic trust

After children judged the two lion puppets, they were asked three epistemic trust questions, each presented on a different slide. In the first slide, an ape was displayed bottom center, and the two lions stood as before on opposite sides on the scene. They had each a small comic strip panel containing the image of either a white ball or a blue ball (Fig. 1). The experimenter told the child that the ape is a new character, that it has a ball and we do not know what color the ball is but we can hear what the two lions have to say. One lion claims that the ball is white, the other that it is blue. Although the child was not provided with direct evidence that the two lions actually saw the ball, the experimenter told that both lions knew the ape well but still gave conflicting testimonies about the color of the ball. The child had to decide whom to trust.

Q1 Color-trust question: "Is the ball blue or white?"

A second question followed on a slide identical to the first one except that the giraffe was now inserted, just above the ape. The giraffe was the puppet that built the tower and interacted with both lions in the character-familiarization movies (one lion destroyed her tower accidentally whereas the other attempted but failed to do so). It was specified that the giraffe had no existing relation with the ape and she should therefore rely on the testimony of one of the two lions. Children were asked to say who the giraffe would believe.

Q2 Giraffe-trust question: "Who does the giraffe believe? Is the ball blue or white, according to the giraffe?"

Last, children saw a slide identical to the first one except that the comic strip panels contained either a smiling green emoji (good) or a frowned red emoji (bad) (the same that were displayed on the boxes in the slides used for the moral judgment questions; see Fig. 1). Children were told that one lion claimed that the ape is good, whereas the other lion said that the ape is bad. It was also clarified that the two lions knew the ape well. The experimenter said, "On their way, the two lions meet the ape, who they both know very well! Now, see what they say. This lion says that the ape is good, whereas this other lion says that the ape is bad." Children were encouraged to decide whom to trust.

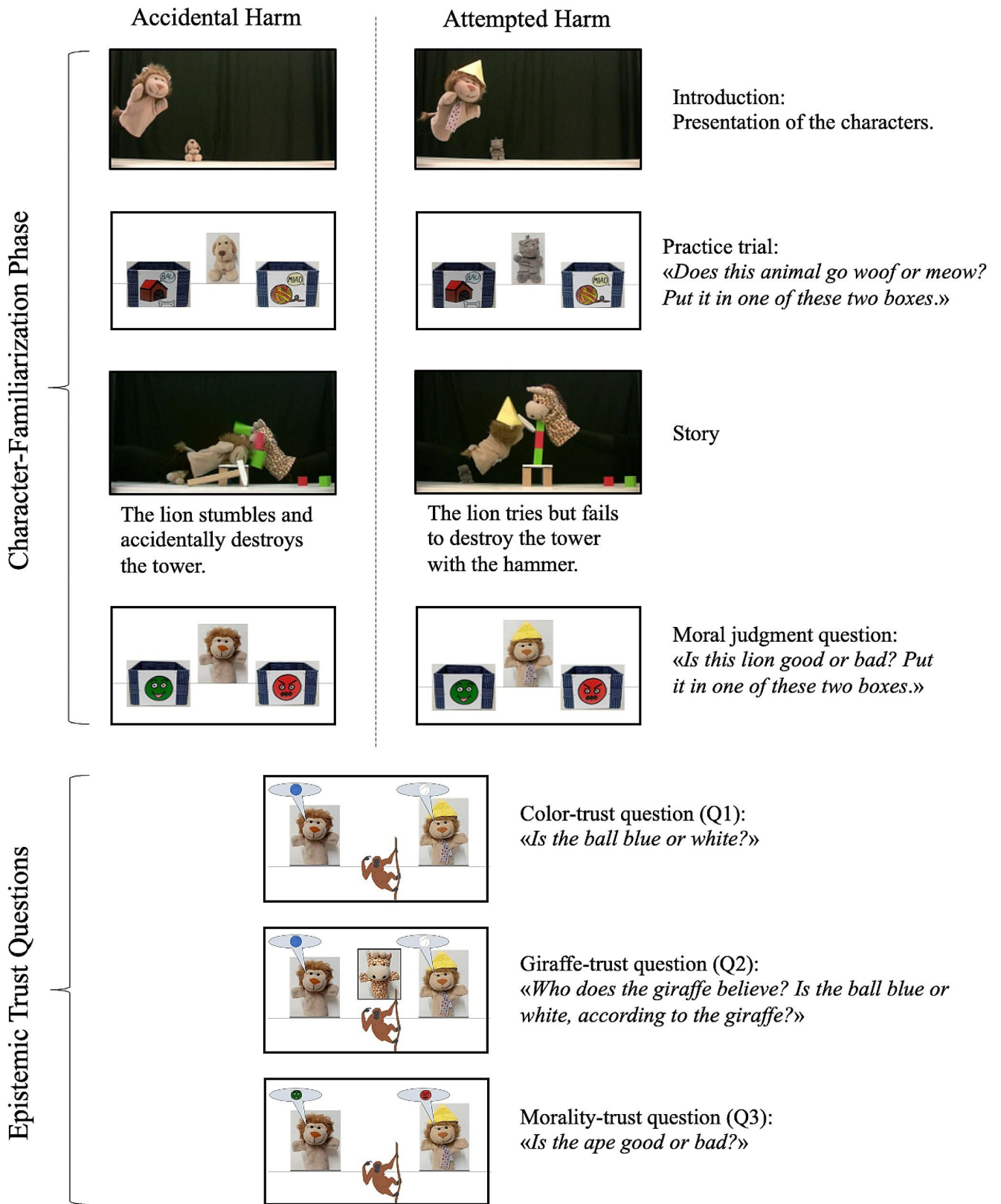


Fig. 1 Schematic depiction of the epistemic trust task. Children saw accidental and attempted harm character-familiarization movies, each preceded by an introductory movie (after which they received a practice trial). After each character-familiarization movie, children received a moral judgment question and had to insert the lion in the

box where the good lions belong (green) or in the box where the bad lions belong (red). Next, children received three questions (Q1-Q3) aimed at assessing which lion they, or the giraffe (Q2), would trust in determining the color of the ape's ball or if the ape was good or bad.

Q3 Morality-trust question: “Is the ape good or bad?”

Thus, with Q1 we tested if children endorsed the testimony of the well-intentioned informant about a physical aspect of an object, with Q2 if they attributed such a tendency to a third agent, and with Q3 if they selectively attended to the well-intentioned agent’s moral testimony. For each question, a score of 1 was attributed if the well-intentioned puppet was trusted and 0 if the ill-intentioned puppet was trusted. Questions were presented in fixed order, from the simplest one, in which children evaluated conflicting testimonies about an object, to the most complex one, in which they evaluated conflicting evaluative statements (likely overcoming their own cognitive biases, a point we shall return to in the discussion).

We counterbalanced: the lions’ position (lion with hat and tie on the left vs. right); the content of the comic strip panels (whether the blue vs. white ball and the green vs. red emoji were on the left vs. right of the slide); the order of options appearing in the questions (e.g., blue vs. white first in Q1, where the first option was always the one on the left on the slide).

Results

Moral judgment

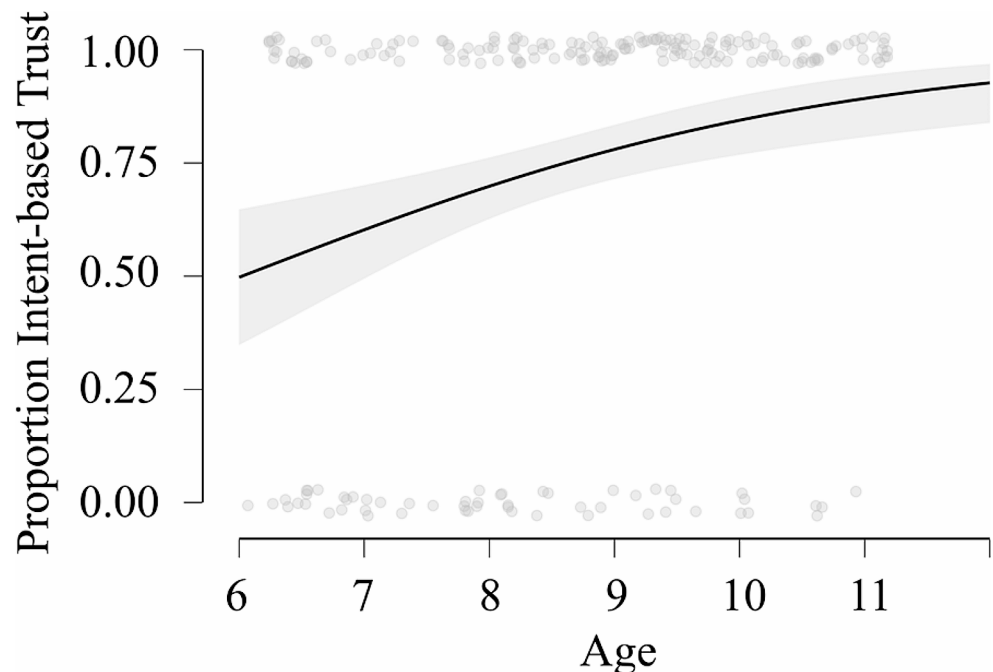
Confirming prior research reporting that by age 5–6 children rely on intentions in moral judgment, across age groups almost all children judged the lion who accidentally harmed as good ($\geq 97\%$), the lion who attempted to harm as bad

($\geq 94\%$), and put the well-intentioned lion in the good box and the ill-intentioned lion in the bad box in response to the two-puppets question ($\geq 94\%$; Table 1).

Epistemic trust

The color-trust question (Q1) was simple to answer for all age groups: most children said that the ball was of the color the well-intentioned lion claimed it to be ($\geq 91\%$; Table 1). Performing a logistic regression analysis on children’s answers with age (continuous) as a predictor confirmed that no significant effect of age was detectable, $\chi^2(1, 218) = 1.16$, Nagelkerke $R = .01$, $p = .283$. A significant effect of age was instead detected in Q2 and Q3. The proportion of children who said that, according to the giraffe (the puppet who built the tower and interacted with the lion puppets in the character-familiarization movies), the ape’s ball was of the color the well-intentioned lion claimed it to be, increased *slightly* with age (from 78% at 6 years to 89% at 10; Table 1), $\chi^2(1, 218) = 4.73$, $R^2 = 0.04$, $p = .033$. However, the majority of 6- and 7-year-olds were already attributing to the giraffe the tendency to endorse the well-intentioned informant’s testimony (78% and 76% respectively vs. 50%, binomial test, $p \leq .005$). Next, the proportion of children who judged the ape according to what the well-intentioned lion said increased with age from 58% at 6 years to 83% at 10 years, $\chi^2(1, 218) = 14.38$, $R^2 = 0.09$, $p < .001$ (Table 1; Fig. 2). Note that for Q3 the proportion of 6- and 7-year-olds who trusted the accidental harm-doer was close to chance level (58% vs. 50%, binomial test, $p \geq .405$). Last, further inspection revealed that most 6- and 7-year-olds in Q3 stated that the ape was good (31/36 at 6 years, 26/33 at 7), which is an

Fig. 2 Logistic regression model depicting the relationship between age and responses to Q3 (1 = trust the accidental harm-doer’s moral testimony). The gray band shows the 95% CI, and datapoints are displayed with the jitter option



important result suggesting that at this age children did not answer randomly but attributed goodness disregarding the informants' testimonies.

Discussion

We tested if 6- to 10-year-olds relied on their own intent-based moral evaluations of the character of two informants to decide which one to trust when the task required to determine the color of an object, who a third party would trust, and the moral quality of an unknown fourth agent. That is, we investigated the development of *intent-based epistemic trust* and found evidence that children's intent-based judgments, well-documented in prior research, do not simply reflect abstract evaluations but also influence children's decisions and learning.

We found that (a) almost all children generated intent-based moral evaluations (they put the well-intentioned puppet in the 'good box' and the ill-intentioned one in the 'bad box'); (b) most children, regardless of age, relied on these evaluations when deciding which of the two informants provided accurate information about the color of a ball; (c) with age, the proportion of children giving an intent-based response in guessing who a third party would trust slightly increased (from 78% at 6 years to 89% at 10), though younger children were already above chance level; (d) the tendency to use informants' past intentions to guide the decision whom to trust about the goodness or badness of a novel agent develops during middle childhood (from 58% at 6 years to 83% at 10).

Result (a) replicates numerous findings showing that children older than 5 generate intent-based moral judgments by condemning failed attempts to harm but not accidental harm (e.g., Margoni & Surian, 2017). Result (a) was not the focus of our study. Result (b) suggests that school-age children are at ease in integrating their own intent-based moral evaluations of informants' characters in their decision whether to trust informants' testimonies, and it is in line with studies showing similar effects already in preschoolers (e.g., Liu et al., 2013). Result (c) hints at a refinement, with development, of the tendency to use intent-based moral evaluations of the characters of informants in foretelling who a third agent will believe, an aspect of children's cognition that has been so far neglected by research.

The greatest novelty of the current study and its true focus is result (d), namely a development in the tendency to trust the well-intentioned informant's moral testimony. Whereas older children trusted the well-intentioned informant, most 6- and 7-year-olds (86% at 6 years and 79% at 7) simply 'trusted' the testimony of the informant who claimed that the novel agent was good, *irrespective* of the

intentions displayed by such informant. This result could be accounted for by different explanations, which are not mutually exclusive.

First, it can be posited that this development is due to a general improvement in using intent-based evaluations in epistemic or selective trust tasks. However, this explanation would not account for why a similar developmental effect was not observed when children were asked to guess the color of the ball. Second, it can be argued that the development unveiled by result (d) is due to the fact that older children get better in understanding testimony, but again we would need to add additional caveats in order to explain why no such development was observed when informants relayed contrasting testimonies about the color of an object. Third, result (d) is *consistent with* the hypothesis that 6- and 7-year-olds had difficulty in suppressing the influence of a default response for which a novel agent is more likely to be good than bad. In turn, this interpretation is *consistent with* a *positivity bias* (Boseovski, 2010), where prior research has shown that such bias has an impact during early to middle childhood on children's use of informant testimony about novel agents' traits. It has been shown that 3- to 7-year-olds trust reliable over unreliable informants more often when informants relayed positive vs. negative information about the personality of a stranger (Boseovski, 2012). At about the same age, children also endorse the testimony of mean vs. nice informants as long as they express positive (vs. negative) judgments (Croce & Boseovski, 2020). Contributing to this bias in evaluating testimony, children, especially until age 7–8, view positive traits as more stable than negative traits (Heyman & Giles, 2004) and require less evidence before attributing positive vs. negative traits (Boseovski & Lee, 2006).

According to this third explanation, what would allow the development unveiled by result (d) is the ability of older children and the inability of younger children to overcome a positivity bias in trait attribution. As of now, this hypothesis should be treated as a speculation deserving attention in future studies. And we cannot but again speculate about what other factors might have contributed to a possible overcoming of a positivity bias in trait attribution in the present study. First, developing mentalistic reasoning and executive functioning skills might have played a role in allowing older but not yet younger children to fully understand that people are not necessarily good by default (Sabbagh & Baldwin, 2001). Second, younger children might have been influenced more than older ones by their real-world experiences of having encountered a majority of good individuals or having lived within the havens of their caring families (for evidence that empirical experience can override information learned through testimony, see Hermansen et al., 2021).

Limitations and future research

A number of limitations can be noted, especially where they can indicate directions for future research. First, we did not test if children were drawing an *explicit* connection between ‘showing a negative intention to smash someone else’s tower’ and ‘being intentionally deceitful’. Future research could investigate whether such an association is made by children. Second, by using a failed attempt scenario to show the presence of a negative intention in one of the two informants, we might have involuntarily depicted that informant as incompetent (the agent tried but failed to do as planned) (see also Vanderbilt et al., 2023). However, if the emergence at 8 years of age of a tendency to perceive incompetent agents as inaccurate were to account for result (d), one would also need to explain why the same development was not observed when children had to guess the color of the ball. This is a problem that the account based on the positivity bias we have outlined above does not have to face. Third, future research could investigate whether our effects were driven by a tendency to trust the well-intentioned informant, to distrust the ill-intentioned informant, or both (Doebel & Koenig, 2013). Fourth, to make our character-familiarization scenarios clear, we used several cues to characterize informants as well- or ill-intentioned (e.g., they verbalized intentions, the accidental harm-doer apologizes and the attempted harm-doer’s ‘victim’ states that the puppet did not succeed in knocking down her tower). Future research could test what cues would suffice to produce the effects we revealed. Fifth, our task might have been too complex for 6- and 7-year-old children, and future research could develop a less cognitively demanding task. Sixth, the order of the trust questions was fixed, and a potential concern might be that this could have caused children to change response criterion from one question to another. However, this possibility would not be consistent with the response pattern we observed; older children were consistently trusting the well-intentioned informant across the trust questions, and younger children were consistent in their responses to the first two questions and they were systematically attributing goodness when replying to the third trust question.

Last, for most of our measures, especially the moral judgment ones, we found low variability in the data. A first possibility is that this fact indeed reflects a real phenomenon in the population, whereas a second possibility is that children’s judgments were the result of following the expectations of relevant adults (e.g., parents and teachers). While the latter remains a possibility, we argue that the former is to prefer, as a number of past studies which used the same or similar stimuli and procedure to the ones we used, have reported evidence of variability in the data when testing younger children (preschoolers), suggesting that a

real change occurs in moral judgment at around the age of 5–6 years (e.g., Cushman et al., 2013; Margoni & Surian, 2017; Nobes et al., 2017). Relatedly, the choice of asking children to judge the moral character of the puppets based on a single action was again consistent with past studies but still required children to make a strong, coarse and unwarranted inference (that is, a single action can hardly tell about the whole character of a person). Thus, a valuable direction for future studies will be to design tasks where characters can be judged based on multiple instances of good or bad behavior, which in turn might help younger children to fully integrate their intent-based moral judgment into their decisions about who to trust.

Conclusions

We investigated the development of intent-based epistemic trust and found that the propensity for taking the moral quality of informants’ past intentions into account to decide whom to trust regarding factual knowledge (the color of a ball) is present at an earlier age than the ability to use such information to attribute moral traits to a novel unknown agent. This work shows that children’s intent-based moral judgment has important downstream effects on children’s decisions and learning outcomes.

Acknowledgements We thank Sara Magri and Alice Agarossi for helping with the data collection, and Tone Kristine Hermansen for providing useful insights during the conceptualization of this research.

Author contributions Conceptualization: FM; Analysis: FM; Supervision: EN; Writing – original draft: FM; Writing – review & editing: FM, EN.

Funding Open access funding provided by University of Stavanger & Stavanger University Hospital.

Data availability This study was not pre-registered; Raw data and study materials, but not study analysis code, are available on Open Science Framework (<https://osf.io/vkf5q/>).

Declarations

Ethical statement We declare that the research project received approval from the University of Milano-Bicocca Ethics Committee (#474) and that parents of all the children involved in this study provided informed consent to participate.

Conflict of interest We have no known conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*, 159–186. <https://doi.org/10.1146/annurev-psych-010213-115033>
- 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*, 1018–1034. <https://doi.org/10.1016/j.cognition.2007.12.008>
- Boseovski, J. J. (2010). Evidence for rose-colored glasses: An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives*, *4*, 212–218. <https://doi.org/10.1111/j.1750-8606.2010.00149.x>
- Boseovski, J. J. (2012). Trust in testimony about strangers: Young children prefer reliable informants who make positive attributions. *Journal of Experimental Child Psychology*, *111*, 543–551. <https://doi.org/10.1016/j.jecp.2011.10.008>
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, *42*, 500–513. <https://doi.org/10.1037/0012-1649.42.3.500>
- Buon, M., Seara-Cardoso, A., & Viding, E. (2016). Why (and how) should we study the interplay between emotional arousal, theory of mind, and inhibitory control to understand moral cognition? *Psychonomic Bulletin & Review*, *23*, 1660–1680. <https://doi.org/10.3758/s13423-016-1042-5>
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Cheung, H., Siu, T. S. C., & Chen, L. (2015). The roles of liar intention, lie content, and theory of mind in children's evaluation of lies. *Journal of Experimental Child Psychology*, *132*, 1–13. <https://doi.org/10.1016/j.jecp.2014.12.002>
- Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*, 360–379. <https://doi.org/10.1111/j.0268-1064.2004.00263.x>
- Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426–437. <https://doi.org/10.1111/j.1467-7687.2008.00792.x>
- Corriveau, K. H., Kinzler, K. D., & Harris, P. L. (2013). Accuracy trumps accent in children's endorsement of object labels. *Developmental Psychology*, *49*, 470–479. <https://doi.org/10.1037/a0030604>
- Croce, R. C., & Boseovski, J. J. (2020). Trait or testimony? Children's preferences for positive informants. *Journal of Experimental Child Psychology*, *190*, 104726. <https://doi.org/10.1016/j.jecp.2019.104726>
- Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, *6*, 97–103. <https://doi.org/10.1016/j.copsyc.2015.06.003>
- Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*, 6–21. <https://doi.org/10.1016/j.cognition.2012.11.008>
- Doebel, S., & Koenig, M. A. (2013). Children's use of moral behavior in selective trust: Discrimination versus learning. *Developmental Psychology*, *49*, 462–469. <https://doi.org/10.1037/a0031595>
- Elashi, F. B., & Mills, C. M. (2014). Do children trust based on group membership or prior accuracy? The role of novel group membership in children's trust decisions. *Journal of Experimental Child Psychology*, *128*, 88–104. <https://doi.org/10.1016/j.jecp.2014.07.003>
- Gillis, R. L., & Nilsen, E. S. (2013). Children's use of information quality to establish speaker preferences. *Developmental Psychology*, *49*, 480–490. <https://doi.org/10.1037/a0029479>
- Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1179–1187. <https://doi.org/10.1098/rstb.2010.0321>
- Henrich, J. (2016). *The secret of our success*. Princeton University Press.
- Hermansen, T. K., Ronfard, S., Harris, P. L., Pons, F., & Zambrana, I. M. (2021). Young children update their trust in an informant's claim when experience tells them otherwise. *Journal of Experimental Child Psychology*, *205*, 105063. <https://doi.org/10.1016/j.jecp.2020.105063>
- Heyman, G. D., & Giles, J. W. (2004). Valence effects in reasoning about evaluative traits. *Merrill-Palmer Quarterly*, *50*, 86–109. <https://doi.org/10.1353/mpq.2004.0004>
- Heyman, G. D., Sritanyaratana, L., & Vanderbilt, K. E. (2013). Young children's trust in overtly misleading advice. *Cognitive Science*, *37*, 646–667. <https://doi.org/10.1111/cogs.12020>
- Hilton, B. C., & Kuhlmeier, V. A. (2019). Intention attribution and the development of moral evaluation. *Frontiers in Psychology*, *9*, 2663. <https://doi.org/10.3389/fpsyg.2018.02663>
- Isella, M., Kanngiesser, P., & Tomasello, M. (2018). Children's selective trust in promises. *Child Development*, *90*, 868–887. <https://doi.org/10.1111/cdev.13105>
- Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*, 106–111. <https://doi.org/10.1111/j.1467-7687.2010.00965.x>
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277. <https://doi.org/10.1111/j.1467-8624.2005.00849.x>
- Koenig, M. A., & Harris, P. L. (2007). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme*, *4*, 264–284. <https://doi.org/10.3366/e1742360007000081>
- 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. <https://doi.org/10.1111/j.0956-7976.2004.00742.x>
- Koenig, M. A., Tiberius, V., & Hamlin, J. K. (2019). Children's judgments of epistemic and moral agents: From situations to intentions. *Perspectives on Psychological Science*, *14*, 344–360. <https://doi.org/10.1177/1745691618805452>
- Li, J., & Tomasello, M. (2018). The development of intention-based sociomoral judgment and distribution behavior from a third-party stance. *Journal of Experimental Child Psychology*, *167*, 78–92. <https://doi.org/10.1016/j.jecp.2017.09.021>
- Li, P. H., Harris, P. L., & Koenig, M. A. (2019). The role of testimony in children's moral decision making: Evidence from China and United States. *Developmental Psychology*, *55*, 2603–2615. <https://doi.org/10.1037/dev0000839>
- Lillard, A. S. (2022). Pretending at hand: How children perceive and process puppets. *Cognitive Development*, *63*, 101202. <https://doi.org/10.1016/j.cogdev.2022.101202>
- Liu, D., Vanderbilt, K. E., & Heyman, G. D. (2013). Selective trust: Children's use of intention and outcome of past testimony. *Developmental Psychology*, *49*, 439–445. <https://doi.org/10.1037/a0031615>
- MacDonald, K., Schug, M., Chase, E., & Barth, H. (2013). My people, right or wrong? Minimal group membership disrupts preschoolers' selective trust. *Cognitive Development*, *28*, 247–259. <https://doi.org/10.1016/j.cogdev.2012.11.001>
- Marble, K. E., & Boseovski, J. J. (2020). Content counts: A trait and moral reasoning framework for children's selective social

- learning. *Advances in Child Development and Behavior*, 58, 95–136. <https://doi.org/10.1016/bs.acdb.2020.01.004>
- Margoni, F., & Surian, L. (2016). Explaining the U-shaped development of intent-based moral judgments. *Frontiers in Psychology*, 7, 219. <https://doi.org/10.3389/fpsyg.2016.00219>
- Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development*, 41, 46–64. <https://doi.org/10.1016/j.cogdev.2016.12.001>
- Margoni, F., & Surian, L. (2020). Conceptual continuity in the development of intent-based moral judgment. *Journal of Experimental Child Psychology*, 194, 104812. <https://doi.org/10.1016/j.jecp.2020.104812>
- Margoni, F., Guglielmetti, G., & Surian, L. (2019). Brief report: Young children with autism can generate intent-based moral judgments. *Journal of Autism and Developmental Disorders*, 49, 5078–5085. <https://doi.org/10.1007/s10803-019-04212-9>
- Margoni, F., Nava, E., & Surian, L. (2022). Do children selectively trust leaders and prosocial agents in an economic exchange? *Developmental Psychology*, 58, 152–160. <https://doi.org/10.1037/dev0001282>
- Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, 131, 716–748. <https://doi.org/10.1037/rev0000450>
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, 112, 367–380. <https://doi.org/10.1016/j.cognition.2009.05.012>
- Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.
- Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General*, 146, 123–133. <https://doi.org/10.1037/xge0000234>
- Nobes, G., Panagiotaki, G., & Bartholomew, K. J. (2016). The influence of intention, outcome and question-wording on children's and adults' moral judgments. *Cognition*, 157, 190–204. <https://doi.org/10.1016/j.cognition.2016.08.019>
- Nobes, G., Panagiotaki, G., & Engelhardt, P. E. (2017). The development of intention-based morality: The influence of intention salience and recency, negligence, and outcome on children's and adults' judgments. *Developmental Psychology*, 53, 1895–1911. <https://doi.org/10.1037/dev0000380>
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, 43, 1216–1226. <https://doi.org/10.1037/0012-1649.43.5.1216>
- Proft, M., & Rakoczy, H. (2018). The ontogeny of intent-based normative judgments. *Developmental Science*, 22, e12728. <https://doi.org/10.1111/desc.12728>
- Rakoczy, H. (2022). Puppet studies present clear and distinct windows into the child's mind. *Cognitive Development*, 61, 101147. <https://doi.org/10.1016/j.cogdev.2021.101147>
- Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development*, 24, 61–69. <https://doi.org/10.1016/j.cogdev.2008.07.004>
- Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.
- Ronfard, S., & Lane, J. D. (2018). Preschoolers continually adjust their epistemic trust based on an informant's ongoing accuracy. *Child Development*, 89, 414–429. <https://doi.org/10.1111/cdev.12720>
- Ronfard, S., & Lane, J. D. (2019). Children's and adults' epistemic trust in and impressions of inaccurate informants. *Journal of Experimental Child Psychology*, 188, 104662. <https://doi.org/10.1016/j.jecp.2019.104662>
- Rosset, E., & Rottman, J. (2014). The big 'whoops!' In the study of intentional behavior: An appeal for a new framework in understanding human actions. *Journal of Cognition and Culture*, 14, 27–39. <https://doi.org/10.1163/15685373-12342108>
- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, 72, 1054–1070. <https://doi.org/10.1111/1467-8624.00334>
- Spelke, E. S. (2022). *What babies know*. Oxford University Press.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., & Origg, G. (2010). Epistemic vigilance. *Mind & Language*, 25, 359–393. <https://doi.org/10.1111/j.1468-0017.2010.01394.x>
- 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. <https://doi.org/10.1111/j.2044-835x.1995.tb00673.x>
- Tong, Y., Wang, F., & Danovitch, J. (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, 23, e12895. <https://doi.org/10.1111/desc.12895>
- Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child Development*, 82, 1372–1380. <https://doi.org/10.1111/j.1467-8624.2011.01629.x>
- Vanderbilt, K. E., Rizzo, M. T., & Blankenship, J. (2023). Preschoolers selectively trust and selectively share with others based on their past accuracy and intentions. *Journal of Experimental Child Psychology*, 228, 105610. <https://doi.org/10.1016/j.jecp.2022.105610>
- Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, 9, 257–270. <https://doi.org/10.1111/j.2044-835x.1991.tb00875.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.