



Social priming enhances young children's trust in robots

Margherita Magnini^{a,*}, Diane Poulin-Dubois^b, Elizabeth J. Goldman^c,
Andrea Luna Tacci^{a,d}, Letizia Aquilino^a, Federico Manzi^{a,d}, Anna E. Baumann^e,
Davide Massaro^a, Antonella Marchetti^a

^a Università Cattolica del Sacro Cuore, Largo Fra Agostino Gemelli, 1, 20123 Milano, MI, Italy

^b Concordia University, 7141 Sherbrooke St W, Montréal, Québec H4B 1R6, Canada

^c Yeshiva University, 245 Lexington Ave, New York, NY 10174, United States

^d IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

^e University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada

ARTICLE INFO

Keywords:

Selective social learning
Epistemic trust
Social robotics
Theory of mind
Priming

ABSTRACT

This study examined preschoolers' selective trust in a robot versus a human informant, and whether social priming may increase trust by making the robot's social characteristics more salient, using the word-learning paradigm. Children also completed the Theory of Mind (ToM) scale, along with an adapted naïve biology task and an Attribution of Mental States questionnaire assessing their beliefs about robots. In Experiment 1, eighty-six Italian children (aged 3 and 5) observed a competent robot (NAO) correctly labeling familiar objects and an incompetent human mislabeling them. The two informants then provided novel labels for unfamiliar objects. Three-year-olds endorsed labels from both informants equally, while five-year-olds preferentially endorsed the labels provided by the robot. ToM related to trust only for judgment scores (which informant was correct), an association largely explained by age. Experiment 2 investigated whether exposure to social priming would increase children's trust. Eighty-five Canadian children (aged 3 and 5) completed the same selective trust procedure, except that they first viewed a social priming video designed to increase the robot's perceived social agency. Both age groups preferred learning new labels from the robot, suggesting that social characteristics support children's trust in robot informants, particularly among younger preschoolers. ToM predicted children's endorsement of the robot's labels, although this effect was largely explained by age. Together, the findings suggest that, by age 5, preschoolers exhibit robust sensitivity to the epistemic characteristics of informants, regardless of their agency. Younger children, in contrast, appear more likely to trust a robot displaying salient social characteristics.

Selective social learning in children

Human culture depends on transmitting information across and within generations, a process shaped by trust in informants (Harris et al., 2012). Children, who regularly encounter new information, learn not only through observation but also from others' reliable testimony (Harris et al., 2018), making the ability to evaluate the trustworthiness of informants crucial. From an early age, children exhibit selectivity in evaluating communicated information, indicating a sensitivity to the reliability of different sources (Ishikawa &

* Corresponding author.

E-mail address: margherita.magnini@unicatt.it (M. Magnini).

Itakura, 2024; Koenig et al., 2004). This ability, known as selective social learning, involves choosing specific informants from whom to acquire knowledge, based on cues such as familiarity, past accuracy, expertise, and group affiliation (Harris et al., 2018; Poulin-Dubois & Brosseau-Liard, 2016). In a recent meta-analysis on selective trust, Tong et al. (2019) identified two key strategies children employ when determining whom to trust. One strategy involves attending to epistemic cues—indicators of an informant's knowledge state, such as past accuracy or demonstrated expertise. For example, preschoolers tend to trust informants who have accurately labeled familiar objects in the past (Koenig et al., 2004). This is particularly advantageous for children with limited prior knowledge, as it helps them obtain reliable information and avoid being misled. The second strategy involves giving more weight to social characteristics, including an informant's age, familiarity, national origin, and benevolent behavior, which reveal social status or relationships (Bascandziew & Harris, 2016; Kinzler et al., 2011; Landrum et al., 2013).

Tong et al. (2019) found that children's reliance on epistemic versus social characteristics shifts with age. When characteristics conflict (e.g., a kind but inaccurate informant versus a mean but accurate one), children begin prioritizing the epistemic ones around age 4. This developmental shift could be explained by improvements in Theory of Mind (ToM), defined as the ability to reason about others' mental states (Wellman, 2014). ToM supports processing informants' intentions and selective learning, as children infer that testimonial accuracy reflects knowledge differences (Crivello et al., 2018; Lucas et al., 2013; Poulin-Dubois & Brosseau-Liard, 2016). Understanding knowledge is key, although the evidence linking ToM and selective trust is mixed. Such link appears to depend on agent type; for example, it is observed with humans but not with puppets (Dragon & Poulin-Dubois, 2023) and weak or absent with robots (Baumann et al., 2023). It is also confounded with age; because ToM and selective trust both develop over the preschool years, the association often disappears once age is controlled for (Baumann et al., 2023). Finally, it varies with how ToM is assessed, emerging with lab-based batteries such as the ToM scale (Brosseau-Liard et al., 2015; Dutemple et al., 2023; Lucas et al., 2013) but not with single tasks or parent-report measures (Pasquini et al., 2007). In sum, until the age of 4, children seem to weigh both types of characteristics equally, suggesting that the ability to prioritize epistemic information develops as children grow older.

Children learning from robots

Today, children have access to a variety of information sources beyond their primary caregivers and other humans, including highly informative technologies (Danovitch and Alzahabi, 2013), raising questions about how children learn from and evaluate the trustworthiness of these sources (Stower et al., 2024).

Humanoid social robots (HSRs), which mimic human appearance and behavior, are especially effective at fostering learning through social interaction. Prior research indicates that children are willing to trust robots and can acquire knowledge from them, similar to how they learn from human informants, by considering the social and epistemic characteristics of the source (Harris et al., 2012). Studies employing a word-learning paradigm (used to study how people learn novel words under different conditions) showed that preschoolers preferred to trust an accurate humanoid NAO robot over an inaccurate one (Brink & Wellman, 2020; Geiskkovitch et al., 2019). Additional evidence from Breazeal et al. (2016), who investigated preschoolers' interactions with two robotic creatures (DragonBot), suggested that contingent behavior enhances both robots' perceived social agency and their effectiveness as learning partners.

Baumann et al. (2023) found that 5-year-olds trust competent robots more than incompetent humans when learning new information, whereas 3-year-olds trust both equally. Li and Yow (2024) observed a developmental shift: Younger children were more likely to accept inaccurate information from humans than robots, suggesting that social characteristics matter more for younger children when informants are non-human. These findings suggest that, when non-human informants are involved, social characteristics become particularly salient for younger children, who tend to accept more inaccurate information from those who display more human-like social characteristics.

Stower et al. (2024) tested 3- to 6-year-olds in three conditions: Reliable human–unreliable robot, unreliable human–reliable robot, and both reliable. Children preferred humans when robots were unreliable, but favored robots otherwise, indicating that accuracy can outweigh agent type. Together, these findings become salient when we consider which characteristics—social or epistemic—most strongly influence trust decisions. This question is closely tied to how children perceive robots and whether attributes such as agency and anthropomorphism play a significant role in shaping trust.

The impact of anthropomorphism

The concepts of animism and anthropomorphism, the perception of inanimate stimuli as alive or sentient, are relevant in this context (Goldman & Poulin-Dubois, 2024). As stated by Goldman and Poulin-Dubois (2024), this perception involves different domains, including biological, psychological, and epistemic trust. If children are more likely to anthropomorphize a robot than adults, could this affect the degree of trust they place in it? From a learning standpoint, anthropomorphizing a robot may lead children to attribute knowledge and competence to it, thereby enhancing its perceived reliability as an informant.

Building on this idea and drawing on the naïve biology task developed by Gottfried and Gelman (2005), multiple studies found that children aged 3 and 5 attribute biological insides to animals and mechanical insides to artifacts, but only from age 5 do they know that robots have mechanical insides (Goldman et al., 2023). To further explore this topic, Baumann et al. (2023) directly compared an inaccurate human informant to an accurate humanoid robot informant (NAO) in Study 1 and to a non-humanoid robot (Cozmo) in Study 2, using the word-learning trust paradigm, to investigate whether robot morphology influences 3- and 5-year-olds' trust decisions. Results showed that 5-year-olds mostly seek and trust the robot's labels, while 3-year-olds alternated between trusting the robot and the human. This suggests that the agency of the robots, not its human-like morphology, guided 3-year-olds' choices.

Interestingly, 5-year-olds endorsed the robot's labels even though they knew the robot had mechanical insides, while 3-year-olds were unsure about them.

In another study, [Baumann et al. \(2024\)](#) investigated how reducing a robot's perceived agency — by removing speech and movement, two rich social cues — affected children's selective trust, diminishing their trust in the robot informant and leading them to trust the inaccurate human informant instead. Using an object search paradigm, 3-year-olds remained ambivalent, showing no clear preference between an accurate non-humanoid robot (Cozmo) and an inaccurate human. These findings suggest that, even without speech, features such as goal-directedness and autonomous movement are sufficient to evoke social affiliation in young children, indicating that 3-year-olds consider both reliability and social characteristics.

In sum, previous work suggests that animism declines with age, although it may persist when salient agency cues are present. Also, children construct trust models of robots, mirroring the trust development mechanisms observed in human–human interaction.

The present study

Children's selective trust seems to be based on two strategies ([Tong et al., 2019](#)): reliance on an informant's epistemic characteristics, such as past accuracy, and on its social characteristics—for a robot, cues such as speech, self-propulsion, and contingency. Because younger children are especially sensitive to informants' social characteristics, enhancing them may raise a robot's perceived social agency and, in turn, increase their willingness to learn from it. Reliance on epistemic characteristics has been shown to be supported by ToM in the case of human informants, which may help children use an informant's prior accuracy (reflecting knowledge) to infer its competence when deciding whom to trust. Importantly, perceiving a robot as a social agent is conceptually distinct from perceiving it as alive (animacy) or as sentient (anthropomorphism), although these constructs may be related.

The current study reports two experiments examining the role of epistemic and social characteristics in 3- and 5-year-olds' selective trust toward a robot versus a human informant. The first and main goal in Experiment 1 was to replicate and extend [Baumann et al. \(2023\)](#) in an Italian sample with an in-person design and direct measures of ToM and anthropomorphism. Testing this sample was intended to assess the robustness of the paradigm when replicated, adding to a growing literature on children's selective trust in robot informants conducted in Canada, the US, Singapore and Australia ([Baumann et al., 2023](#); [Brink & Wellman, 2020](#); [Li & Yow, 2024](#); [Stower et al., 2024](#)). Baumann and colleagues showed that 3-year-old Canadian children endorsed labels from both the robot and the human, while 5-year-olds favored the competent robot—even knowing that it was mechanical—a robust test of epistemic trust. Children's trust was assessed using the selective trust task ([Koenig et al., 2004](#)) with a competent robot and an incompetent human. We hypothesized that this age-related pattern would replicate.

The second goal was to investigate the impact of social characteristics on children's trust in robots versus humans. While prior work ([Baumann et al., 2024](#)) examined the effects of removing such characteristics to determine whether trust decreases, in Experiment 2, the robot's social characteristics were enhanced by having children primed with a video in which the robot engaged in a contingent, turn-taking conversation with a person. Previous findings also suggest that even minimal cues—such as pointing—can signal agency and lead children to perceive the robot as a social agent, though not necessarily enough to elicit full trust. Experiment 2 examines whether children's trust differs in a context intended to make the robot's social characteristics more salient. We hypothesized that this manipulation would increase children's trust in the robotic informant, particularly among 3-year-olds, who are most sensitive to informants' social characteristics.

A third goal—pursued across both experiments—was to clarify the inconsistent link between selective trust and ToM discussed above. In contrast to previous studies with robots, which tested ToM with a parent-report measure or a single task, we measured ToM with a direct, multi-task scale ([Wellman & Liu, 2004](#)), expecting this more sensitive measure to reveal an association with selective trust toward a non-human informant, which we then tested before and after controlling for age.

Lastly, to examine how children perceive the robot, we administered a naïve biology task (a test of animacy) that allowed us to determine whether children's trust in robots relates to an accurate judgment of the robot as an artefact. We also assessed children's attribution of mental states to the robot and the human using an adapted Attribution of Mental States (AMS) questionnaire ([Miraglia et al., 2023](#)), which indexes anthropomorphism.

Experiment 1

Method

Participants

An a priori G*Power 3.1 analysis ([Faul et al., 2007](#)) was conducted to determine the appropriate sample size for a 2 (age; between-subjects) $\times$ 3 (trial type; within-subjects) mixed-design ANOVA. The goal was to obtain .80 power to detect a medium effect size of .25 at the standard .05 alpha error probability. Based on the between-subjects effect, the least sensitive in this design, the analysis revealed that a minimum sample size of 43 per group was required.

Children were recruited from three schools located in the regions of Piemonte and Lombardia, Italy, and tested in Italian. A total of 100 children were tested, with the sample size determined by school availability and parental consents. Fourteen children were excluded from the analyses after data collection due to the child's distraction during the testing session ($n = 5$), language-related difficulties ($n = 3$), and familiarity with humanoid robots ($n = 6$). The final sample consisted of 86 children, distributed across two age groups: 3-year-olds ($n = 46$, $M_{age} = 41.24$ months, $SD = 3.45$, $N_{male} = 23$) and 5-year-olds ($n = 40$, $M_{age} = 65.20$ months, $SD = 4.25$, $N_{male} = 23$). With this final sample, power for the between-subjects effect was .756 under the two-level model specification (see

Results).

The parent/guardian completed a survey about their child's demographics and prior robot exposure to ensure the child had no previous robot knowledge. Of the entire sample, 6 families (6 %) reported that their children interacted daily with humanoid robots (i. e., during play). In the final sample, participants were mainly European (91.9 %), with the remaining participants identifying as African (2.3 %), Indigenous (1.2 %), Other (2.3 %), or Unsure (1.2 %). The University's Ethics Committee approved the present study.

Materials

The experiment used a NAO, an autonomous, programmable humanoid robot (23 in, 5.6 kg, 25 degrees of freedom). A MacBook Air (13-inch, 2017) equipped with PowerPoint was used to administer both the selective trust task and naïve biology task, projecting the visual stimuli. Additional materials for the selective trust task included various toys (a car, ball, cup, twine, plastic bulb, and funnel).

ToM-scale materials comprised nine figurines, three dolls, and two boxes. Additional figurines of humans and a robot were included in the materials for the AMS questionnaire. For a more complete list and pictures of all the materials, see Supplementary Materials.

Procedure

Participants were recruited by contacting preschools in Italy. Following approval from school coordinators and teachers, recruitment materials—including a study presentation letter, informed consent form, and a socio-demographic questionnaire—were distributed to parents of children aged 3 or 5 years. Families received further information and could ask questions in online meetings.

Unlike Baumann et al. (2023), testing was conducted in person to reduce distractions and ensure protocol validity, taking place during school hours in a quiet room provided by each participating school. Children were tested individually in the presence of two experimenters. Each session lasted approximately 20 min. The testing sessions were video-recorded and stored on a university server. The tasks were administered in the following order: ToM scale, naïve biology task (counterbalanced before or after the selective trust task to control for potential robot-trial effects on children's agency), AMS questionnaire, and selective trust task.

Theory of Mind. To measure ToM skills, the Wellman & Liu (2004) ToM scale was administered. It consists of seven items of increasing difficulty: Diverse desires, diverse beliefs, knowledge access, contents false belief, explicit false belief, belief emotion, and real apparent emotion. Only the first 5 items were administered because the last two are intended for children aged 6 (our sample consisted of 3- and 5-year-olds). One image was adapted to better fit the Italian cultural context. In the diverse beliefs task, a canteen replaced a garage because "garage" can be ambiguous in Italian, whereas a canteen is familiar and easily recognizable to most Italian children. Children receive a point for each item, as each is scored pass/fail (range = 0–5). ToM scores were correlated with children's trust decisions. Beyond such correlations and descriptive statistics, findings of the ToM scale are reported in Supplementary Materials.

Attribution of Mental States. An adapted version of the AMS questionnaire (Miraglia et al., 2023) was administered to participants before the selective trust task. While the original questionnaire, validated for adults, comprises 23 items, the present adaptation was specifically designed for use with children and consisted of 8 items grouped into four categories: Epistemic (think, remember), sensory (see, hear), emotional (be happy, be sad), and volitional (dream, tell a lie). Children responded to each item using a three-point scale (Di Dio et al., 2020): No (0 points), Yes, a bit (1 point), or Yes, a lot (2 points). The task was administered under two conditions: Human or Robot. The total score was calculated as the sum of all answers (range = 0–16).

Naïve Biology. The naïve biology task was adapted from Gottfried and Gelman (2005). The task began with two training trials. Each training trial featured an image of a familiar furniture item (i.e., a fridge or closet) missing a central piece, denoted by a white rectangle. Next to the familiar item were two items children could choose from. One option plausibly belonged inside the target item (e.g., food for the fridge) and the other did not (e.g., clothing). During the training trials, the experimenter demonstrated which inside was correct, and then the child had to choose between the two options independently.

During the 9 test trials, children were shown four unfamiliar animals (i.e., ibek, pacarana, tapir, cavy), four unfamiliar artifacts (i. e., intercom, coffee machine, voice recorder, electric razor), and a picture of the robot NAO (see Fig. 1 for an example of a trial).

For each trial, one of four biological insides (i.e., muscle, lungs, heart, bone) and one of four mechanical insides (i.e., gears, circuit,

Naïve biology task – Example of the robot trial

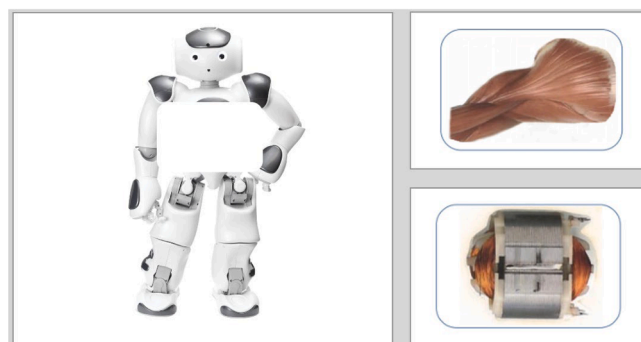


Fig. 1. Naïve biology task – Example of the robot trial.

batteries, wires) were presented as a pair. The experimenter asked the child which of the two options should go inside. Children responded verbally, indicating which option they thought belonged inside (top or bottom), or pointed at it. The test trials, internal insides paired with each unfamiliar image, the order of the unfamiliar images, and whether the biological or mechanical inside was on the top or the bottom of the screen were counterbalanced, resulting in four conditions. Children received a point each time they properly assigned the correct inside to the target picture (range: 0 – 9). [Gottfried and Gelman \(2005\)](#) verified that the objects featured in the task are novel to children.

Selective Trust Task. The selective trust task was adapted by [Baumann et al. \(2023\)](#) from [Koenig et al.'s \(2004\)](#) paradigm. Children watched pre-recorded videos in which a robot (NAO) and a female adult (Francesca) labeled different objects. Both informants were introduced as either a robot or a human, along with their names and presented as the experimenter's friends. An example video is available in the OSF repository.

Introductory video. The experimenter told the children they had brought toys to play with and showed an introductory video in which the robot and the human introduced themselves by name, each saying it was excited to play a game.

Familiarization trials. In three familiarization trials, NAO always labeled familiar objects correctly, whereas the human incorrectly (see [Table 1](#)). After each trial, the child was asked to endorse one of the labels provided by the informants (endorse question: "Can you tell me what this is called?"). Following the three trials, the child was asked to indicate which informant had provided correct or incorrect information (judgment question: "My friends just told you a lot of things, did either of them say something [right/wrong]?"), allowing children to verbally express which informant they deemed reliable.

Test trials. In three test trials, NAO and the human labeled unfamiliar objects (a blue cylinder, a white rubber bulb, and a red silicone funnel). Before watching videos of NAO and the human labeling novel objects, the children were asked whether they knew what each object was called. If a child gave a label, the experimenter replied: "That's a good guess, but I don't think that's what this is called. Let's see if our friends can help us figure it out". The child was then prompted: "Who would you like to ask how this is called?" (ask question), used to identify any bias toward the robot and to examine whether the familiarization trials rendered NAO reliable. After that, the informants labeled the objects with different nonsense labels, and the child was asked to choose and name the object using one of the informants' labels (endorse questions), requiring the child to endorse the more competent one, providing a direct measure of epistemic trust. After the test trials, the child indicated which informant provided correct or incorrect labels (test judgment question). Across four versions, object order was fixed, while speaker order, speaker position, and the judgment question were counterbalanced. Children received 1 point for each ask, endorse, or judgment response favoring NAO.

Results

All participants were live coded by a second experimenter present during the testing sessions. If a child switched their answers, the child's final response was used. A small number of participants did not answer certain questions, resulting in some empty cells ($n = 22$), and an additional eight children did not complete the AMS questionnaire, due to experimenter error or the child's refusal to answer. These empty cells in the datasheet were treated as missing data points and left blank. All data were checked for normality. If a deviation from normality was found, appropriate corrections were applied, and nonparametric tests were run. All analyses were carried out using IBM SPSS Statistics 30.0.0.0 (172) and JASP 0.95.2. A complete report of all descriptive statistics is provided in [Table 2](#). Results of the residual normality checks and partial correlation for both Experiments are reported in the Supplementary Materials.

Table 1
Familiar and unfamiliar objects and labels used in the selective trust task.

Familiar Objects	Correct Label – NAO	Incorrect Label – human
	Car	Book
	Ball	Shoe
	Cup	Dog
Unfamiliar Objects	Correct Label – NAO	Incorrect Label – human
	Toma	Mido
	Fep	Dax
	Bosa	Dawnoo

Selective trust analysis

Familiarization trials

In the endorse familiarization trials, children responded with the correct label 96.5 % of the time, confirming that children were familiar with these items.

Test trials – Ask and endorse questions

An independent-samples *t*-test showed no performance difference between children who offered a name for an object and those who did not across any of the trials (all *p*'s > .21).

Chance analyses. Chance analyses showed that both 3- and 5-year-olds children showed a preference for the robot to help them find the label of the novel objects as shown by the ask questions (3-year-olds: $M = 2.08$, $SD = 1.07$; $t(45) = 3.71$, $p < .001$, $d = 0.54$, 95 % CI [0.26, 0.90]; 5-year-olds: $M = 2.52$, $SD = 0.64$; $t(39) = 10.12$, $p < .001$, $d = 1.60$, 95 % CI [1.12, 2.06]). A different pattern emerged for the endorse trials, with 3-year-olds performing at chance, endorsing equally the human and the robot informant, while 5-year-olds prioritizing competency and learning more from the robot than the person (3-year-olds: $M = 1.80$, $SD = 1.16$; $t(45) = 1.76$, $p = .08$, $d = 0.26$, 95 % CI [-0.03, 0.55]; 5-year-olds: $M = 2.25$, $SD = 0.95$; $t(39) = 4.97$, $p < .001$, $d = 0.78$, 95 % CI [0.42, 1.13]).

Linear Mixed Model. A linear mixed-effects model fitted by REML was run on the number of times children chose the robot in the selective trust task across the ask and endorse trials. Ask and endorse trials were modelled jointly as they share the same response scale (0–3), whereas judgment trials operate on a different scale (0–2) and were therefore analyzed separately. The gender of the child (male or female), the age of the child (3- or 5-year-old), and trial type (ask or endorse) were entered as fixed effects. As repeated observations were nested within participants, a by-participant random intercept was included since each child was asked all questions. The model yielded a significant effect between age groups ($b = 0.43$, 95 % CI [0.07, 0.80], $SE = .18$, $p = .019$), gender ($b = 0.43$, 95 % CI [0.08, 0.78], $SE = .18$, $p = .018$), and trial type ($b = 0.27$, 95 % CI [0.08, 0.47], $SE = .09$, $p = .005$), with no age $\times$ trial type interaction ($p = .96$). Estimated marginal means indicated that 5-year-olds (EMM = 2.42, $SE = .13$) chose the robot more than 3-year-olds (EMM = 1.98, $SE = .12$), male children (EMM = 2.42, $SE = .13$) picked the robot more than female children (EMM = 1.99, $SE = .12$), and children asked more for the name of an object to the robot (EMM = 2.34, $SE = .10$) than they endorsed its label (EMM = 2.06, $SE = .10$). This indicates no significant age-group differences in response patterns across trial types, but a difference in overall tendency to choose the robot informant, qualifying the pattern suggested by the chance analyses above.

Judgment trials

We separately examined whether the familiarization and test judgment trials were above chance level. This result was confirmed for both age groups. A binomial test showed that 34 out of 40 5-year-olds (85 %) in the test trial selected the reliable informant, which was above chance ($p < .001$), as did 39 out of 40 (98 %) in the familiarization trial. Also, 3-year-olds succeeded in both the test trial (31 out of 46, 67 %, $p = .02$) and the familiarization trial (33 out of 46, 72 %, $p = .005$).

An independent-samples *t*-test was conducted to compare the combined judgment scores between 3-year-olds and 5-year-olds. There was a significant difference in scores between the two age groups, $t(76.67) = -3.42$, $p = .001$, $d = -0.72$, 95 % CI [-1.15, -0.28]. Five-year-olds ($M = 1.83$, $SD = 0.45$) scored significantly higher than three-year-olds ($M = 1.39$, $SD = 0.71$).

Naïve biology analysis

Chance analyses indicated that both 3- ($M = 2.51$, $SD = 1.16$), $t(42) = 2.89$, $p = .006$, $d = 0.44$, 95 % CI [0.12, 0.75]) and 5-year-old children ($M = 3.58$, $SD = 0.87$), $t(39) = 11.40$, $p < .001$, $d = 1.80$, 95 % CI [1.29, 2.30]) demonstrated an understanding that animals possess biological insides. In contrast, performance on the artifacts varied with age. Three-year-olds did not differ from chance in identifying mechanical insides ($M = 2.31$, $SD = 1.04$), $t(41) = 1.91$, $p = .06$, $d = .29$, 95 % CI [-0.02, 0.60]. However, 5-year-olds performed significantly above chance ($M = 3.58$, $SD = 0.71$), $t(39) = 13.99$, $p < .001$, $d = 2.21$, 95 % CI [1.62, 2.78]. Independent samples *t*-test revealed that 5-year-olds outperformed 3-year-olds both in the animal and artifacts conditions (Animal: $t(77.66) = -4.73$, $p < .001$, $d = -1.02$, 95 % CI [-1.48, -.56]; Mechanical: $t(72.5) = 6.43$, $p < .001$, $d = -1.40$, 95 % CI [-1.88, -.91]).

Table 2

Descriptive statistics for all tasks by age group (3- and 5-year-olds).

Task	3-year-olds		5-year-olds	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Theory of Mind scale	1.63	.927	3.47	1.43
Naive Biology Machines	2.31	1.04	3.58	.712
Naive Biology Animals	2.51	1.16	3.57	.874
Naive Biology Robot	.58	.499	.65	.483
AMS Human	9.86	3.88	12.00	4.27
AMS Robot	8.74	4.62	6.12	4.75
Ask	2.08	1.07	2.52	.640
Endorse	1.80	1.16	2.25	.954
Judgment combined	1.39	.714	1.82	.446

Note. *M* = Mean, *SD*=Standard Deviation.

For the robot trial, binomial tests on mechanical choices showed that both age groups performed at chance, attributing biological and mechanical insides equally (3-year-olds: 26 out of 45, 58 %, $p = .371$; 5-year-olds: 26 out of 40, 65 %, $p = .081$). No difference between the two age groups was found in the robot condition, as well as in performance between the children who performed the task before or after the selective trust task (all p 's $> .2$).

Attribution of mental states analysis

Statistical analyses indicated that both 3- and 5-year-olds children attributed mental states to humans above chance (one sample t -test, 3-year-olds: $M = 9.86$ ($SD = 3.88$), $t(43) = 3.18$, $p = .003$, $d = .48$, 95 % CI [0.16, 0.78]; Wilcoxon one-sample test, 5-year-olds: $W = 397.00$, $N = 31$, $p < .001$). In contrast, developmental differences emerged in the attribution of mental states to robots. Three-year-olds performed at chance in this condition ($W = 528.00$, $N = 46$, $p = .34$), suggesting potential confusion or inconsistent understanding in this domain. Conversely, five-year-olds attributed mental states to robots at rates significantly different from chance ($W = 97.50$, $N = 32$, $p = .02$), indicating a developing recognition that robots do not possess human-like mental states.

Independent samples t -test revealed that 5-year-olds outperformed 3-year-olds in both conditions (Human: $t(73) = -2.25$, $p = .02$, $d = -0.52$, 95 % CI [-.99, -.05]; Robot: $t(76) = 2.43$, $p = .01$, $d = 0.55$, 95 % CI [0.09, 1.01]).

Correlational analyses

To examine a potential link between ToM and the other variables, we computed a series of correlation coefficients. ToM scores were significantly related to the attribution of mental states to robots ($r = -.30$, $p = .007$), indicating that children with more advanced ToM skills are less likely to attribute mental states to robots. Across the selective trust task, the correlation between ToM and trust was significant only for total judgment scores ($r = .28$, $p = .008$), indicating that children who performed better on ToM were more consistent in judging the robot as the correct informant. All correlations are reported in Table 3.

Partial correlations controlling for age were conducted to examine the associations between ToM and the other variables. These correlations were no longer significant, suggesting that the originally observed associations were largely accounted for by age-related variance.

Discussion

The primary aim of this experiment was to examine whether 3- and 5-year-olds prefer learning from a competent robot over an incompetent human, extending Baumann et al.'s (2023) findings to a sample of Italian children so that the robustness of the paradigm could be confirmed in a new population and with in person testing. When comparing our results with those reported by Baumann et al. (2023). Specifically, 3-year-olds showed no clear preference, endorsing labels provided by both the human and the robot at comparable rates. In contrast, 5-year-olds, as expected, selectively endorsed the label provided by the competent robot over that of the incompetent human. However, this pattern was not confirmed by a formal interaction test in the LMM, which showed an overall quantitative advantage for 5-year-olds instead. This age-related shift is broadly consistent with the developmental trajectory reported by Baumann et al. (2023) with Canadian children, suggesting that selective trust based on epistemic reliability extends to non-human agents and becomes more consistent by age 5. Additionally, gender differences were found in children's choices of whom to seek new words from, with female participants preferring to ask the human informant more often than male participants. This may be because the human informant in the selective trust task was female, leading female participants to rely more on social characteristics, like gender group membership, when making decisions.

Regarding the robot's inanimate status, the AMS confirmed that 5-year-olds knew that NAO was not a sentient agent (they did not attribute mental states to it), yet they still chose to learn from it. In contrast, 3-year-olds were ambivalent about attributing biological attributes and mental states to the robot and about whom to endorse during the test trials. Regarding the attribution of internal properties to animals and artifacts, our findings replicate previous work.

One result that differs from the original study concerns the attribution of biological versus mechanical internal properties to the

Table 3
Bivariate Correlations between ToM score and task performance – Experiment 1.

	1	2	3	4	5	6	7	8
1. Theory of Mind	--							
2. Naïve Biology – Machines	.421**	--						
3. Naïve Biology – Animals	.262*	.513**	--					
4. Naïve Biology – Robot	-.011	.109	.189	--				
5. AMS – Human	.161	.243*	.151	.142	--			
6. AMS – Robot	-.304**	-.211	-.184	.185	.405**	--		
7. Ask	.194	.326**	.358**	.019	.098	.037	--	
8. Endorse	.183	.344**	.096	-.094	.187	.036	.619**	--
9. Judgment combined	.283**	.218*	.062	.053	.145	-.059	.244*	.378**

Note. Values are Pearson correlation coefficients (r). $N = 86$. ** represent significance at the 0.01 level (2-tailed); * represent significance at the 0.05 level (2-tailed).

robot. While Baumann et al. (2023) found that 5-year-olds recognized that the robot possesses mechanical insides, Italian children performed at chance, suggesting that they did not clearly distinguish what goes inside a robot. This discrepancy may be due to methodological differences. In the present work, the robot trial was presented only once, whereas in Baumann et al. (2023), children completed two trials. To address this limitation, we improved the design of the naïve biology task in Experiment 2.

Finally, we found a ToM–trust correlation only in the case of judgment trials; however, this association did not survive a control for age. This suggests that ToM may play a role but is not the primary factor in the shift from social to epistemic trust when an informant is non-human. Children may rely more on social than epistemic characteristics in such cases. To explore this further, we conducted a follow-up experiment using the same procedure with an added robot–child social priming manipulation. By portraying the robot as both competent and socially engaging, we examined whether enhanced social characteristics would influence younger children's selective trust.

Experiment 2

The second experiment investigated whether exposure to social priming would increase children's willingness to learn from a non-human informant. Specifically, whether such priming could reduce the ambivalence previously observed in 3-year-olds, who tend to learn equally from an incompetent human and a competent non-human informant when deciding whom to learn new words from. Prior research suggests that the presence of agency cues in robots—such as gaze and contingent responsiveness—can enhance young children's imitation and engagement with non-human agents (Breazeal et al., 2016; Manzi et al., 2020). However, other studies have failed to find such effects on word mapping (O'Connell et al., 2009), leaving the relation between agency cues and learning unclear. To address these mixed findings, Experiment 2 aimed to determine whether designing a contingent interaction between a robot and a human, thereby enhancing the robot's perceived social agency, would improve children's learning from a competent non-human informant. We hypothesized that exposure to the priming video would enhance 3-year-old children's learning from the competent non-human informant, while 5-year-olds would continue to be sensitive to competence, choosing the robot.

Method

Participants

A total of 90 children were tested. Five children were excluded from the analyses after data collection due to incomplete task participation ($n = 3$) or familiarity with humanoid robots ($n = 2$). The final sample consisted of 85 children living in a large Canadian metropolitan city, tested to match with the population in which the original paradigm was developed, while introducing the priming within the same cultural context. Children were distributed across two age groups: 3-year-old ($n = 43$, $M_{age} = 41.02$ months, $SD = 3.34$, $N_{male} = 24$) and 5-year-old children ($n = 42$, $M_{age} = 66.69$ months, $SD = 2.93$, $N_{male} = 23$). The a priori power analysis mirrored that of Experiment 1 with the final sample ($N = 85$) and the two-level model specification, power for the between-subjects effect was .748 for the assumed $f = .25$.

Children were recruited from an existing database of participants and tested in a small university laboratory room. The testing session was conducted in either English ($n = 61$) or French ($n = 24$), reflecting the dominant languages spoken by the local population. As compensation, parents received a \$25 gift card to a local bookstore and a certificate of merit for their participation.

Only neurotypical children were included in the sample. The parent/guardian completed a survey about their child's demographics and prior robot exposure to ensure the child had no previous robot knowledge. Two families (2.3 %) reported that their children interacted daily with humanoid robots (i.e., during play activities) and were excluded from the final sample. In the final sample, participants were mainly European (48.3 %), with the rest of the sample identifying as Mixed (26.4 %), Asian (12.6 %), Middle Eastern (4.6 %), Latin, Central or South American (2.3 %), African (2.3 %), and prefer not to answer (3.4 %). The sample was generally composed of families of high socioeconomic status, with 63.5 % earning more than \$100,000 per year and another 17.7 % earning more than \$50,000 per year. The University's Ethics Committee approved the present study.

Materials

The materials used in Experiment 2 were identical to those employed in Experiment 1, with two modifications. First, children watched the priming video and were tested on the selective trust task via a larger screen (Samsung 32" 720p HD LED) connected to an iPad running PowerPoint. Second, we improved the design of the naïve biology task (Baumann et al., 2024). The main change in Experiment 2 involved manipulating the perceived sociability of the robot. Specifically, children were primed with a 90-second video before the trust task. Because Experiment 1 revealed no differences as a function of the order of administration of the naïve biology task, tasks were administered in the following fixed order: ToM scale, naïve biology task, AMS questionnaire, and selective trust task.

Naïve biology task

Revisions were made to better understand how children perceive the robot and human agents in the selective trust task. The task used two laminated cards, each showing the human or robot agent with a missing internal part. Each agent was presented across four trials, and an additional interview question was asked: "Is the human/robot alive?" (alive question). Four experimental conditions were created for this task, with both the order of the alive question and the agent type (human/robot) counterbalanced. For the alive question, two additional laminated cards were used, showing the human and robot without missing parts. Finally, small, laminated cards were employed to depict the mechanical and biological internal parts.

Social priming

Experiment 2 aimed to enhance the robot's social characteristics in order to test its impact on children's trust. Following Baumann et al. (2024), who reduced such characteristics by removing speech and autonomous movement, the priming video did the reverse: it enhanced the social characteristics (speech, self-propulsion, contingency) to increase the robot's perceived social agency.

The video depicted the robot and the human actor who appeared in the selective trust task engaged in a conversation about animals. The dialogue was constructed by analyzing human-to-human communication patterns and modeling the robot's interaction on established social norms. Specifically, we drew on principles of sequential organization (in which each contribution displays understanding of the prior partner's contribution), turn-taking (minimizing gaps and overlaps in speech), and contingency (e.g., greetings followed by reciprocal responses), tailoring the script to the recipient (Pelikan & Broth, 2016). In addition to verbal interaction, we emphasized the importance of nonverbal communication by focusing on gestures. Prior research indicates that social perception is shaped approximately 7 % by verbal cues, 38 % by vocal tone, and 55 % by bodily cues, underscoring the central role of nonverbal behaviors in evaluations of social agents (Chidambaram et al., 2012). To this end, we incorporated design recommendations from prior work (Gockley et al., 2005; Chidambaram et al., 2012), focusing on gesture production and contingency behavior, as suggested by Breazeal et al. (2016). The robot's repertoire included deictic gestures (e.g., pointing to itself or to the participant accordingly), beat gestures (rhythmic hand movements to emphasize speech), and iconic gestures (depicting objects or events through representational movement). These behaviors were synchronized with speech to reinforce the impression of social competence. Scripts are provided in the Supplementary Materials.

Procedure

Parents were contacted by email and telephone and, before the testing session, were briefed on the purpose of the study and provided written informed consent on behalf of their children. After the session, parents were debriefed about its goals and could ask questions. The testing session was recorded, and the videos were stored on a secure university server.

Results

All participants were coded by two coders who watched recordings of the testing sessions. If a child switched their answers, the child's final response was used. A small number of participants did not answer certain questions, resulting in some empty cells ($n = 6$ trials) due to experimenter error or the child's refusal to answer. These empty cells in the datasheet were treated as missing data points and left blank. All data were checked for normality. If a deviation from normality was found, appropriate corrections were applied, and nonparametric tests were run. All analyses were carried out using IBM SPSS Statistics (2024) and JASP 0.95.2. All descriptive statistics are reported in Table 4.

Selective trust analysis

Familiarization trials

When asked what the familiar items were called, children responded with the correct label 98.1 % of the time, confirming that children were familiar with these items.

Test trials – Ask and endorse questions

An independent sample t -test showed no performance difference between children who offered a name for an object versus those who did not, on any of the trials (all p 's > 0.12).

Chance analyses. Chance analyses showed that both 3- and 5-year-olds children showed a preference for the robot to help them find the label of the novel objects, the ask questions (3-year-olds: $M = 2.21$, $SD = 1.12$; $t(42) = 4.13$, $p < .001$, $d = 0.63$, 95 % CI [0.30, 0.95]; 5-year-olds: $M = 2.55$, $SD = 0.80$; $t(41) = 8.46$, $p < .001$, $d = 1.30$, 95 % CI [0.88, 1.71]). The same pattern emerged for the endorse trials, with 3- and 5-year-olds learning the label for the unfamiliar objects from the robot than from the person (3-year-olds: $M = 1.95$, $SD = 0.92$; $t(42) = 3.21$, $p = .003$, $d = 0.49$, 95 % CI [0.17, 0.80]; 5-year-olds: $M = 2.50$, $SD = 0.80$; $t(41) = 8.06$, $p < .001$, $d = 1.24$, 95 % CI [0.83, 1.64]).

Table 4

Descriptive statistics for all tasks by age group (3- and 5-year-olds).

Task	3-year-olds		5-year-olds	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Theory of Mind scale	1.53	1.24	3.24	1.30
Naive Biology Humans	2.40	1.05	3.88	.395
Naive Biology Robot	2.02	1.18	3.31	1.33
AMS Human	10.07	5.21	12.29	3.81
AMS Robot	8.00	5.64	6.12	4.86
Ask	2.21	1.12	2.55	.803
Endorse	1.95	.925	2.50	.804
Judgment combined	1.40	.760	1.86	.354

Note. *M* = Mean, *SD*=Standard Deviation.

Linear Mixed Model. The same linear mixed-effects model as in Experiment 1 was run, with the addition of the language of testing (French or English) as a control fixed effect, given that children were tested in either language. The model yielded significant main effects of age ($b = 0.44$, 95 % CI [0.11, 0.77], $SE = .16$, $p = .009$) and gender ($b = 0.33$, 95 % CI [0.00, 0.66], $SE = .16$, $p = .050$), with no significant effects of trial type, language, or the age $\times$ trial type interaction (all p 's $> .16$). Estimated marginal means indicated that 5-year-olds ($EMM = 2.48$, $SE = .12$) chose the robot more than 3-year-olds ($EMM = 2.04$, $SE = .12$), and male children ($EMM = 2.42$, $SE = .11$) more than female children ($EMM = 2.09$, $SE = .13$).

Judgment trials

We separately examined whether the familiarization and test judgment trials were above chance level. This result is confirmed for the 5-year-olds, where a binomial test showed that in the test trial 36 out of 42 children (86 %) selected the competent informant, scoring above chance level, $p < .001$; as well as in the familiarization trial (42 out of 42, 100 %). In contrast, 3-year-olds performed at chance level (27 out of 43, 63 %; $p = .12$) in the test trial. Still, as expected, they succeeded in the familiarization trial (33 out of 43, 76 %). An independent samples t -test on the combined judgment scores revealed a significant difference in mean performance between the two age groups, $t(59,72) = -3.60$, $p < .001$, $d = -0.77$, 95 % CI [-1.21, -0.33], with 5-year-olds outperforming 3-year-olds.

Naïve biology analysis

Human and robot trials

Chance analyses indicated that both 3- and 5-year-olds children understand that the human possessed biological insides (3-year-olds: $M = 2.40$, $SD = 1.05$, $t(42) = 2.47$, $p = .01$, $d = 0.37$, 95 % CI [0.65, 0.68]; 5-year-olds: $M = 3.88$, $SD = 0.39$, $t(41) = 30.84$, $p < .001$, $d = 4.75$, 95 % CI [3.68, 5.82]). In contrast, performance on the robot trials varied with age. Three-year-olds did not differ from chance in identifying the mechanical insides of robots ($M = 2.02$, $SD = 1.18$), $t(42) = 0.12$, $p = .89$, $d = 0.02$, 95 % CI [-0.27, 0.31], suggesting confusion or lack of knowledge in this domain. However, 5-year-olds performed significantly above chance ($M = 3.31$, $SD = 1.33$, $t(41) = 6.36$, $p < .001$, $d = 0.98$, 95 % CI [0.60, 1.34]), indicating a clearer understanding of the mechanical insides of robots.

Independent samples t -test revealed that in both conditions 5-year-olds outperformed 3-year-olds (Human: $t(53,91) = -8.67$, $p < .001$, $d = -1.86$, 95 % CI [-2.37, -1.35]; Robot: $t(83) = -4.70$, $p < .001$, $d = -1.02$, 95 % CI [-1.47, -0.56]).

Alive questions

When asked, “Do you think that the human is alive?” both age groups respond “yes” above chance level (3-year-olds: 37 out of 43, 86 %; 5-year-olds: 34 out of 41, 81 %). Similar results emerged on the robot trial (“Do you think that the robot is alive?”); 37 out of 43 3-year-olds (86 %) responded “yes”, which was significantly above chance level, all p 's $< .001$. In the robot trial, 28 out of 42 5-year-olds (67 %) responded yes ($p = .04$); even though the result is significant, there is a tendency towards greater awareness that robots are not alive compared to the 3-year-olds ($p < .001$).

The two age groups differed significantly only in the robot condition: 5-year-olds recognize that the robot is not alive more than 3-year-olds ($\chi^2(1, N = 85) = 4.43$, $p = .03$), whereas the human condition showed no group difference ($p = .5$).

Attribution of mental states analysis

A Wilcoxon signed-rank test indicated that both 3- and 5-year-olds attributed mental states to humans at rates significantly greater than chance (3-year-olds: $W = 609.00$, $N = 43$, $p = .02$; 5-year-olds: $W = 791.50$, $N = 42$, $p = .01$). Developmental differences emerged in the AMS to robots. Three-year-olds perform at chance level in this condition ($W = 450.50$, $N = 43$, $p = .99$), suggesting potential confusion or inconsistent understanding in this domain. Conversely, 5-year-olds attribute fewer mental states to robots ($W = 230.50$, $N = 42$, $p = .01$), indicating a developing recognition that robots lack mental states.

An independent-samples t -test revealed that only in the human condition did 5-year-olds outperform 3-year-olds ($t(76.95) = -2.24$, $p = .02$, $d = -0.48$, 95 % CI [-.91, -.05]), whereas in the robot condition, there was no group difference ($p = .10$).

Table 5

Bivariate Correlations between ToM score and task performance – Experiment 2.

	1	2	3	4	5	6	7
1. Theory of Mind	--						
2. Naïve Biology – Human	.328**	--					
3. Naïve Biology – Robot	.292**	.409**	--				
4. AMS – Human	.161	.194	.192	--			
5. AMS – Robot	-.113	-.123	-.225*	.401**	--		
6. Ask	.221*	.054	.076	.195	.108	--	
7. Endorse	.231*	.091	.200	.092	.004	.438**	--
8. Judgment combined	.307**	.329**	.280**	.221*	.026	.361**	.376**

Note. Values are Pearson correlation coefficients (r). $N = 85$. ** represent significance at the 0.01 level (2-tailed); * represent significance at the 0.05 level (2-tailed).

Correlational analyses

As we wanted to examine a potential link between ToM and the other variables, we computed a series of correlation coefficients. ToM was positively associated with the naïve biology task in both conditions (Human: $r = 0.32$, $p = .002$; Robot: $r = 0.29$, $p = .007$). ToM was also positively associated with all three measures of trust: Endorse test trials ($r = 0.23$, $p = .03$), ask trials ($r = 0.21$, $p = .04$), and the combined judgment score trials ($r = 0.30$, $p = .004$). No significant correlations were found between ToM and the attribution of mental states to both humans and robots. All correlations are reported in full in Table 5. As in Experiment 1, partial correlations controlling for age were no longer significant, suggesting that the previously observed association between ToM and the other variables were explained by age.

Theory of Mind and trust

To better investigate the correlation between ToM and trust, a simple linear regression was conducted to examine whether children's ToM predicted performance in the trust trials. The overall regression model, $F(1, 83) = 4.69$, $p = .033$, indicates that ToM was a significant predictor of the endorsement question (i.e., endorsing the label of the accurate robot informant). The model explained approximately 5.3 % of the variance ($R^2 = .053$, adjusted $R^2 = .042$). The unstandardized regression coefficient indicated that higher ToM scores were associated with higher scores on the endorse trials ($B = 0.137$, $SE = 0.063$, $\beta = .231$, $t(83) = 2.16$, $p = .033$).

Selective trust across experiments

Because the two experiments differed in the nationality of the children, the following comparison characterizes the pattern across the two samples rather than testing the priming manipulation per se, and observed differences must be interpreted with caution. We conducted independent-samples *t*-tests for the 3- and 5-year-old age groups to examine potential differences between experiments on the trust measures. Analyses for each age group indicated no significant differences between conditions (all p 's $> .1$). Notably, as shown in Fig. 2, a modest endorsement effect emerged only when age groups were combined. Specifically, children in Experiment 1 (who were not exposed to the priming video) scored significantly lower than those in Experiment 2 (who viewed the priming video), $t(169) = -2.01$, $p = .04$, $d = -0.30$, 95 % CI $[-0.60, -0.005]$. The ask and the judgment test trial, as expected, were not significant (all p 's $> .55$). This effect should be interpreted with caution, as it emerged only with age groups combined.

Discussion

Experiment 2 examined whether enhancing the robot's social characteristics would increase trust and offset the age effects observed in previous studies. Both age groups performed well in the selective trust task, seeking information and endorsing the competent robot. Notably, the effect was more evident in endorsement trials, which more directly capture children's willingness to accept information from the robot. While chance analyses suggest increased endorsement of the robot, particularly for 3-year-olds, the comparison across experiments did not reveal significant age-specific effects of the priming manipulation; accordingly, this pattern

Bar plot showing ask, endorse, and judgment test trial scores across the two experiments

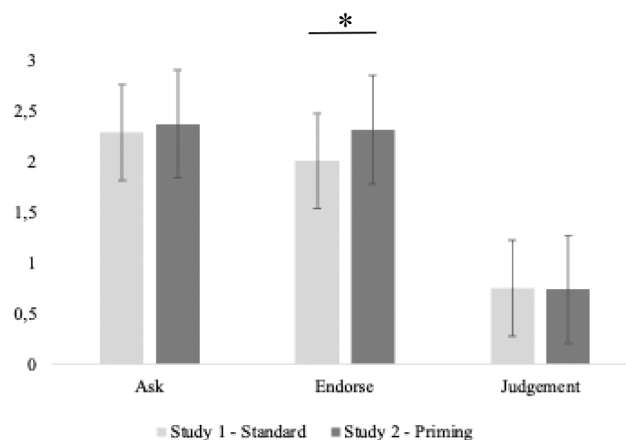


Fig. 2. Bar plot showing ask, endorse, and judgment test trial scores across the two experiments. Note. Ask and endorse scores range from 0 to 3, while the judgment test trial scores range from 0 to 1. Error bars represent standard error. The between columns * represent significance in between groups below $p < .05$.

provides limited evidence for age-related differences in responses to social priming.

In Experiment 2, only the 5-year-olds consistently recognized the robot as a competent informant during the judgment test trial. One possible explanation for this discrepancy lies in the cognitive demands of the judgment question. This trial requires children to recall multiple statements, evaluate their accuracy, and attribute correctness to a specific informant. Younger children may find this metacognitive process challenging, particularly when novel objects are involved, as uncertainty about object properties can amplify difficulty. In contrast, 5-year-olds—who typically demonstrate more advanced epistemic reasoning—may have been better equipped to infer the robot's competence under these conditions.

Regarding the robot's inanimate status, 5-year-olds knew that NAO was inanimate (they did not attribute mental states to it and correctly recognized its mechanical insides), yet they still chose to learn from it, relying on its epistemic characteristics. This interpretation is only partly mirrored by the direct *"Is the robot alive?"* question: although a majority of 5-year-olds still answered "yes", they were more likely than 3-year-olds to recognize that the robot is not alive. In contrast, 3-year-olds were ambivalent about the robot's animacy status regarding biological attributes and mental states. For humans, our findings replicate and extend previous work, showing that both 3- and 5-year-olds know that humans possess biological insides and attribute mental states to them. In Experiment 2, ToM was associated with all three trust measures, but these associations did not survive a control for age, providing limited evidence that mentalizing independently predicts trust and suggesting a broader developmental change instead.

In sum, our findings are broadly consistent with the hypothesis that social characteristics may play a role in preschoolers' epistemic trust toward non-human informants. Younger children's greater willingness to rely on the robot to learn new words is consistent with the priming having made its social characteristics more salient, rather than with a demonstrated change in how children perceived it. However, because no manipulation check was included, the underlying mechanism remains unclear; future work could assess the effects of priming by directly measuring children's perceptions, such as friendship or preference toward the robot.

General discussion

The main goal of the current set of studies was to examine the role of the informants' social and epistemic characteristics in 3- and 5-year-old children's selective trust, using a word-learning paradigm across two experimental conditions. In the first, standard condition, a competent robot was pitted against an incompetent human, allowing us to investigate whether children's learning choices were guided by epistemic trust (Baumann et al., 2023). The second condition was designed to enhance the robot's social characteristics and determine whether this manipulation, combined with competence, would lead 3-year-olds to make a clear choice about from whom to learn new words.

Consistent with our first hypothesis, 5-year-olds consistently preferred the competent robot over the incompetent human in both studies. In contrast, 3-year-olds trusted both informants equally in Experiment 1 but favored the robot in Experiment 2. Because the two experiments involved samples drawn from different cultural groups, we interpret this pattern cautiously, as preliminary evidence that younger children may be more willing to learn from a robot when its social characteristics are made salient.

Portraying the robot as knowledgeable and socially engaging may reduce younger children's reluctance to learn from a non-human informant. However, we cannot determine whether the priming video altered children's perceived social agency, as the AMS and naïve biology tasks always preceded the priming; Experiment 2 should therefore be interpreted as a test of children's trust in a context designed to increase the salience of the robot's social characteristics, rather than as a demonstration that their perceptions were changed. Testing this context is nonetheless theoretically motivated: enhancing a robot's social characteristics may raise its perceived social agency and, in younger children, their willingness to learn from it—the inverse of Baumann et al. (2024), who reduced these characteristics. The cross-experiment pattern is also less likely to reflect cultural differences alone, given that Italian preschoolers form trust relationships with both human and robotic partners (Di Dio et al., 2020), and that Experiment 1 reproduced the North American developmental pattern (Baumann et al., 2023) in an Italian sample, with unprimed 3-year-olds showing ambivalence in both Western samples.

Importantly, across both experiments, the developmental differences in trust were most evident on the endorsement trials. Of the three trust measures (ask, endorse, and judge), the endorse question is the only one that captures genuine trust and learning: The ask trials may instead reflect a novelty bias toward the robot rather than true learning from the informants. Likewise, the judgment test trial may be challenging for children, as they cannot always determine whether an informant was correct or incorrect. In contrast, the endorse question offers less ambiguity and fewer alternative interpretations, making it a more reliable indicator of selective trust.

Beyond trust, 3-year-olds remained uncertain about the robot's internal properties (biological or mechanical) and about its mental states across both experiments; as these measures were collected before the priming, they reflect baseline conceptions rather than any change produced by the manipulation.

In addition to our main analysis, we aimed to determine whether ToM abilities could predict epistemic trust. Across both experiments, ToM provided limited independent evidence of predicting trust once age was controlled, suggesting that the age-related shift toward the competent robot reflects broader developmental changes rather than a specific effect of ToM. Previous studies with non-human agents suggest that this association may emerge when interacting with humanoid robots, such as NAO (Study 1; Baumann et al., 2023), whereas no such link has been observed with non-humanoid robots like Cozmo (Study 2; Baumann et al., 2023) or when the robot's perceived agency is reduced (Baumann et al., 2024). When such a link is observed, its presence may depend on the informant's morphology and agency cues, suggesting that they trigger knowledge attribution. Building on this idea, Goldman et al. (2025) propose that one possible explanation for why children attribute more epistemic properties to humans than to robots is given by the context of the task in which the robot is presented. When stimuli are presented within familiar, everyday situations, children are more likely to treat robots as depictions of social agents than when they are presented without a context, such as the AMS questionnaire. This raises

the possibility that a human-like context—such as a robot-person interaction, as in our Experiment 2— could make it easier for children to anthropomorphize robots and consequently perceive them as competent. Testing this directly, with a manipulation check, is an important direction for future work.

Importantly, the present work offers several novel contributions to the existing literature. Firstly, we replicated the main results of a previous experiment within an Italian context and using an in-person design, thereby confirming the robustness of the epistemic trust paradigm. Secondly, all administered tasks were designed to provide a comprehensive view of children's perceptions of robots and how these relate to their overall cognitive development. Indeed, directly assessing naïve biology, mental state attribution, and ToM provides valuable insight into how children trust and learn from social robots. Finally, to our knowledge, no studies have employed a protocol directly comparing a competent robot and an incompetent human while specifically enhancing the robot's social characteristics to examine its influence on 3- and 5-year-olds' trust. Previous research reduced the robot's agency to favor the human informant (Baumann et al., 2024). However, those findings suggested that even minimal cues—such as pointing—can serve as sufficiently strong indicators of agency to make children perceive the robot as a social agent, though not necessarily enough to elicit complete trust. For this reason, the current study adds clarity regarding whether a context designed to make the robot's social characteristics more salient influences children's epistemic trust in robots.

Limitations and future directions

The current work has several limitations that should be addressed in future studies. To more accurately assess the real impact of social priming on anthropomorphism, a follow-up study could measure the AMS questionnaire and naïve biology after the priming and include explicit measures of social affiliation (e.g., friendship, liking, or partner-preference judgments). Including manipulation checks would help clarify the specific effects of priming on children's perceived social agency of the robot, including whether changes in trust arise from increased attribution of mental states or greater perceived social closeness. One limitation is that the priming manipulation was introduced in a different cultural group from Experiment 1, so that sample nationality and manipulation are fully confounded; we therefore cannot determine whether the differences observed across the two experiments reflect the priming, cultural factors, or both. However, because the Italian children behaved like other samples in the trust literature a purely cultural explanation seems less likely. A definitive test would require a fully crossed culture $\times$ priming design together with a direct manipulation check of children's perception of the robot, which we identify as priorities for future work.

Furthermore, to better elucidate the link between ToM and selective trust, it would be useful to incorporate ToM items that directly assess epistemic competencies. For instance, the Comprehensive Assessment of Theory of Mind (CAT; Heise & Bowman, 2025), which includes the Knowledge Expertise item, could provide a more nuanced examination of ToM abilities in the domain of epistemic understanding. Such measures would contribute to a deeper understanding of how cognitive and social factors jointly shape children's selective trust in robotic informants.

In conclusion, the present work confirms a developmental shift in epistemic trust between the ages of 3 and 5. By age 5, children recognize the competence of both human and non-human informants. It also suggests that social characteristics may shape younger children's trust, as enhancing a robot's social characteristics appeared to facilitate children's willingness to engage with and learn from the robot. These findings have important implications for educational robotics. Understanding how children attribute social and epistemic qualities to robots can help educators and designers create social robots that foster effective learning and positive engagement in early childhood.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author(s) used ChatGPT in order to reframe portions of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

CRediT authorship contribution statement

Margherita Magnini: Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Diane Poulin-Dubois:** Supervision, Resources, Conceptualization. **Elizabeth J. Goldman:** Writing – original draft. **Andrea Luna Tacci:** Investigation. **Letizia Aquilino:** Investigation. **Federico Manzi:** Writing – original draft. **Anna E. Baumann:** Writing – original draft. **Davide Massaro:** Supervision, Conceptualization. **Antonella Marchetti:** Supervision, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank all members of the Research Center on Theory of Mind at the Università Cattolica del Sacro Cuore and all members of the Cognitive and Language Development Laboratory at Concordia University, especially Lorraine Foley, for her valuable contribution in testing. We also want to acknowledge the Scuola Materna Ronzoni Silva, Scuola dell'Infanzia Fratelli Russi, and Scuola dell'Infanzia Sacra Famiglia for their valuable collaboration, as well as all the families that participated in the study. Margherita Magnini was supported by a PhD scholarship (no. 7279, 39th cycle) awarded within the framework of the Italian National Recovery and Resilience Plan (NRRP/PNRR), Mission 4, Component 1, Investment 4.1, funded by the European Union – NextGenerationEU under Ministerial Decree no. 118/2023 (CUP J53C23001160001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jecp.2026.106618>.

Data availability

OSF Link Provided.

References

- Bascandziev, I., & Harris, P. L. (2016). The beautiful and the accurate: Are children's selective trust decisions biased?. ISSN 0022-0965 *Journal of Experimental Child Psychology*, 152, 92–105. <https://doi.org/10.1016/j.jecp.2016.06.017>.
- Baumann, A. E., Goldman, E. J., Meltzer, A., & Poulin-Dubois, D. (2023). People do not always know best: Preschoolers' trust in social robots. *Journal of Cognition and Development*, 24(4), 535–562. <https://doi.org/10.1080/15248372.2023.2178435>
- Baumann, A. E., Goldman, E. J., Cobos, M.-G.-M., & Poulin-Dubois, D. (2024). Do preschoolers trust a competent robot pointer? *Journal of Experimental Child Psychology*, 238, Article 105783. <https://doi.org/10.1016/j.jecp.2023.105783>
- Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science*, 8(2), 481–491. <https://doi.org/10.1111/tops.12192>
- Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, 56(7), 1268–1277. <https://doi.org/10.1037/dev0000884>
- Brousseau-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>
- Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In Proceedings of the 7th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12) (pp. 293–300). Association for Computing Machinery. <https://doi.org/10.1145/2157689.2157798>.
- Crivello, C., Phillips, S., & Poulin-Dubois, D. (2018). Selective social learning in infancy: Looking for mechanisms. *Developmental Science*, 21(3), Article e12592. <https://doi.org/10.1111/desc.12592>
- Danovitch, J. H., & Alzhabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, 14(3), 499–513. <https://doi.org/10.1080/15248372.2012.689391>
- Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I trust you? From child–robot interaction to trusting relationships. *Frontiers in Psychology*, 11, 469. <https://doi.org/10.3389/fpsyg.2020.00469>
- Dragon, M., & Poulin-Dubois, D. (2023). To copy or not to copy: A comparison of selective trust and overimitation in young children. *Cognitive Development*, 66, Article 101316. <https://doi.org/10.1016/j.cogdev.2023.101316>
- Dutemple, E., Hakimi, H., & Poulin-Dubois, D. (2023). Do I know what they know? Linking metacognition, theory of mind, and selective social learning. *Journal of Experimental Child Psychology*, 227, Article 105572. <https://doi.org/10.1016/j.jecp.2022.105572>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Geiskovitch, D. Y., Thiessen, R., Young, J. E., & Glenwright, M. R. (2019). What? That's not a chair! how robot informational errors affect children's trust towards robots. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 48–56). <https://doi.org/10.1109/HRI.2019.8673024>
- Gockley, R., & Broz, Frank & Forlizzi, J. & Michalowski, M. & Mundell, A. & Rosenthal, Stephanie & Sellner, B. & Simmons, Reid & Snipes, K. & Schultz, Alan & Wang, Jue. (2005). Designing robots for long-term social interaction. IEEE Conference on Intelligent Robots and Systems (IROS 2005). 1338 - 1343. 10.1109/IROS.2005.1545303.
- Goldman, E. J., & Poulin-Dubois, D. (2024). Children's anthropomorphism of inanimate agents. *Wiley Interdisciplinary Reviews: Cognitive Science*, 15(4), e1676.
- Goldman, E. J., Baumann, A. E., & Poulin-Dubois, D. (2023). Preschoolers' anthropomorphizing of robots: Do human-like properties matter? *Frontiers in Psychology*, 13, Article 1102370. <https://doi.org/10.3389/fpsyg.2022.1102370>
- Goldman, E. J., Baumann, A.-E., Paré, L., Beaudoin, J., & Poulin-Dubois, D. (2025). Children's attribution of mental states to humans and social robots assessed with the Theory of Mind Scale. *Scientific Reports*, 15, 16357. <https://doi.org/10.1038/s41598-025-96229-7>
- Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks: The role of insides and immanence. *Cognitive Development*, 20(1), 137–158. <https://doi.org/10.1016/j.cogdev.2004.07.003>
- Harris, P. L., Corriveau, K. H., Pasquini, E. S., Koenig, M., Fusaro, M., & Clément, F. (2012). Credulity and the development of selective trust in early childhood. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 193–212). Oxford University Press.
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, 69(1), 251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>
- Heise, M. J., & Bowman, L. C. (2025). The comprehensive assessment of theory of mind (CAT): A novel measure of 3- to 8-year-old children's theory of mind and an evaluation of mental-state scaling. *Child development*, 96(5), 1787–1806. <https://doi.org/10.1111/cdev.14263>
- Ishikawa, M., & Itakura, S. (2024). The development of social learning: From pedagogical cues to selective learning. *Frontiers in Psychology*, 15, Article 1466618. <https://doi.org/10.3389/fpsyg.2024.1466618>
- Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native accented speakers. *Developmental Science*, 14(1), 106–111. <https://doi.org/10.1111/j.1467-7687.2010.00965.x>

- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, 15(10), 694–698. <https://doi.org/10.1111/j.0956-7976.2004.00742.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>
- Li, X., & Yow, W. Q. (2024). Younger, not older, children trust an inaccurate human informant more than an inaccurate robot informant. *Child development*, 95(3), 988–1000. <https://doi.org/10.1111/cdev.14048>
- Lucas, A. J., Lewis, C., Pala, F. C., Wong, K., & Berridge, D. (2013). Social-cognitive processes in preschoolers' selective trust: Three cultures compared. *Developmental Psychology*, 49(3), 579–590. <https://doi.org/10.1037/a0029864>
- Manzi, F., Ishikawa, M., Di Dio, C., et al. (2020). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, 10. <https://doi.org/10.1038/s41598-020-69140-6>
- Miraglia, L., Peretti, G., Manzi, F., Di Dio, C., Massaro, D., & Marchetti, A. (2023). Development and validation of the Attribution of Mental States Questionnaire (AMS-Q): A reference tool for assessing anthropomorphism. *Frontiers in Psychology*, 14, Article 999921. <https://doi.org/10.3389/fpsyg.2023.999921>
- O'Connell, L., Poulin-Dubois, D., Demke, T., & Guay, A. (2009). Can infants use a nonhuman agent's gaze direction to establish word-object relations? *Journal of Cognition and Development*, 11(1), 81–98. <https://doi.org/10.1080/15250000902994073>
- 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. <https://doi.org/10.1037/0012-1649.43.5.1216>
- Pelikan, H. R. M., & Broth, M. (2016). Why that Nao? How humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 4921–4932). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858478>
- Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The developmental origins of selective social learning. *Current Directions in Psychological Science*, 25(1), 60–64. <https://doi.org/10.1177/0963721415613962>
- Stower, R., Kappas, A., & Sommer, K. (2024). When is it right for a robot to be wrong? Children trust a robot over a human in a selective trust task. *Computers in Human Behavior*, 157, Article 108229. <https://doi.org/10.1016/j.chb.2024.108229>
- Tong, Y., Wang, F., & Danovitch, J. (2019). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, 23(2), Article e12895. <https://doi.org/10.1111/desc.12895>
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523–541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>