Cognition 249 (2024) 105814

Contents lists available at ScienceDirect

Cognition

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

School-age children are more skeptical of inaccurate robots than adults

Teresa Flanagan [a,*], Nicholas C. Georgiou [b], Brian Scassellati [b], Tamar Kushnir [a]

[a] Department of Psychology & Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27701, United States of America
[b] Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06511, United States of America

ARTICLE INFO

Keywords:
Cognitive development
Educational robotics
Selective trust
Human-robot interaction

ABSTRACT

We expect children to learn new words, skills, and ideas from various technologies. When learning from humans, children prefer people who are reliable and trustworthy, yet children also forgive people's occasional mistakes. Are the dynamics of children learning from technologies, which can also be unreliable, similar to learning from humans? We tackle this question by focusing on early childhood, an age at which children are expected to master foundational academic skills. In this project, 168 4–7-year-old children (Study 1) and 168 adults (Study 2) played a word-guessing game with either a human or robot. The partner first gave a sequence of correct answers, but then followed this with a sequence of wrong answers, with a reaction following each one. Reactions varied by condition, either expressing an accident, an accident marked with an apology, or an unhelpful intention. We found that older children were less trusting than both younger children and adults and were even more skeptical after errors. Trust decreased most rapidly when errors were intentional, but only children (and especially older children) outright rejected help from intentionally unhelpful partners. As an exception to this general trend, older children maintained their trust for longer when a robot (but not a human) apologized for its mistake. Our work suggests that educational technology design cannot be one size fits all but rather must account for developmental changes in children's learning goals.

Every day, we put our trust in technologies – we ask smart speakers, like Amazon Alexa, what the temperature is, we use apps to learn new languages, we seek help from chatbots, like ChatGPT, to write our code, emails, and even papers. When we are adults, these instances of trust may seem like small additions to what we already know about the world, but what if we were expected to trust technologies as we are forming the foundations of our knowledge? In this project, we address this question by investigating whether 4- to 7-year-old children, and by comparison adults, trust an interactive technology that starts out accurate and helpful but becomes inaccurate and unhelpful mid-way through a collaborative word learning game.

Educational technologies for young children became popular in the 1990s with TV programs like *Baby Einstein* and have drastically increased in multiple mediums in the 30 years since. Though parents and educators were initially excited for the possibilities of broadening access to learning opportunities for young children, research warned that relying too heavily on TV and other technologies (including eBooks and some tablet applications) did more harm than good (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004). For example, numerous studies on the "video deficit effect" (Anderson & Pempek, 2005; Barr, 2010; Strouse & Samson, 2021; Troseth, 2003; Troseth & DeLoache, 1998) show that children under 3-years-old do not encode simple events (new words, locations of hidden objects) from video, but can easily learn when the same events are shown to them live. Even though video deficits decline with age, research continues to show that best practice for media use is as a supplement to social interactions with adults, rather than as stand-alone experiences (Myers, Crawford, Murphy, Aka-Ezoua, & Felix, 2018; Nielsen, Simcock, & Jenkins, 2008; Strouse & Ganea, 2017; Strouse, Troseth, O'Doherty, & Saylor, 2018). The message is clear: *social learning* is best for children, and technology is only effective when it is interactive (i.e., can contingently communicate either verbally or nonverbally), rather than passively used.

Enter robots – interactive, social technologies with the potential to transform children's education by virtue of their form and function (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018). Robots are an interesting case for educational technologies because they are readily perceived by children (and even sometimes by adults) to be agentic – not exactly human, but not exactly objects either (Bernstein & Crowley, 2008; Brink, Gray, & Wellman, 2019; Chernyak & Gary, 2016; Fiala, Arico, & Nichols, 2014; Flanagan, Wong, & Kushnir, 2023; Fussell,

* Corresponding author.
E-mail address: tmf34@duke.edu (T. Flanagan).

https://doi.org/10.1016/j.cognition.2024.105814
Received 22 September 2023; Received in revised form 10 May 2024; Accepted 13 May 2024
Available online 18 May 2024

Kiesler, Setlock, & Yew, 2008; Jipson & Gelman, 2007; Kahn et al., 2012; Kahn et al., 2012). Promisingly, new work shows that robots make better teachers for young children than TV and other non-interactive media. For example, toddlers' vocabulary skills improve when taught by a social robot (Movellan, Eckhardt, Virnes, & Rodriguez, 2009), preschoolers learn novel words and facts taught by robots (Breazeal et al., 2016; Brink & Wellman, 2020; Kory Westlund et al., 2017; Li & Yow, 2023), and children even engage with robots in uniquely human forms of social learning, such as imitation (though not as much as they are willing to do so for humans, Sommer et al., 2020).

Thus, by mimicking the qualities of human social teachers, robots (and, more recently, AI applications without human-like form but with interactive, agentic qualities; see Araujo, 2018; Girouard-Hallam & Danovitch, 2022; Kasneci et al., 2023) have the potential to transform education in early childhood. But for this potential to be realized, we need a better understanding of how to design systems that match the expectations of young social learners, who have evolved learning mechanisms responsive to uniquely human social cues and forms of communication (Tomasello, 2019).

In this paper, we explore whether one such characteristic of human social learning applies to robots: that learning is built on a foundation of trust which can be maintained even after occasional mistakes. In any give social interaction where information is shared – whether it be between speakers and listeners or between teachers and learners – there is a tacit acknowledgement that communicators will be truthful and maximally informative, and thus learners (or listeners) can trust the quality of the information they receive (Bonawitz et al., 2011; Grice, 1989; Gweon, 2021; Landrum, Eaves, & Shafto, 2015). While young children are inclined to trust information from others (Jaswal, Croft, Setia, & Cole, 2010), they also have an emerging ability to selectively block information from teachers who violate this acknowledgment (e.g., by being repeatedly inaccurate or uncertain – Harris & Corriveau, 2011; Koenig & Harris, 2005a; Koenig & Woodward, 2010; Sobel & Kushnir, 2013).

Equally important for human social learning is the ability to understand the communicative signals of occasional mistakes *without* losing trust. From the second year of life, young children distinguish between errors that are intentional and accidental (Behne, Carpenter, Call, & Tomasello, 2005; Carpenter, Call, & Tomasello, 2005; Gergely, Bekkering, & Király, 2002). As such, preschoolers and elementary-aged children will forgive an accidental transgression (Amir, Ahl, Parsons, & McAuliffe, 2021; Chernyak & Gary, 2016; McElroy, Kelsey, Oostenbroek, & Vaish, 2023). By 4-years-old, children will still learn from a teacher who is occasionally wrong, as long as the teacher has also been occasionally right (Pasquini, Corriveau, Koenig, & Harris, 2007; Ronfard & Lane, 2018). Preschoolers will also continue to trust a person that admits ignorance about some things but shows confidence in their knowledge of other things (Kushnir & Koenig, 2017; Kushnir, Vredenburgh, & Schneider, 2013). Preschoolers also understand that displays of remorse after a mistake, such as an apology, can be an indication that the transgressor feels guilty and wants to rectify the mistake (Smith, Chen, & Harris, 2010; Smith & Harris, 2012). Accordingly, 4-year-olds prefer and forgive a person that gives an apology after a transgression (Oostenbroek & Vaish, 2019; Vaish, Missana, & Tomasello, 2011).

However, children are not forgiving of all mistakes. For example, 3–6-year-olds will promote punishment on those that intentionally cause harm (Chernyak & Gary, 2016; McElroy et al., 2023). Preschoolers also distrust and negatively evaluate people that have a history of harmful behavior (Doebel & Koenig, 2013) and people that intentionally share false information (D'Esterre, Rizzo, & Killen, 2019; Rizzo, Li, Burkholder, & Killen, 2019). Children also do not treat all apologies as the same, but this seems to change with age. Specifically, 5–6-year-olds are less forgiving of a person that repeatedly gives the same apology for the same offense (Waddington, Jensen, & Köymen, 2022) and 7–9-year-olds distinguish between willing and coerced apologies (Smith, Anderson, & Straussberger, 2018). Together this work suggests that young children expect and prefer their human teachers to be truthful and well-intentioned, while also understanding that humans cannot be perfect.

Can lessons from this body of work be applied to learning from robots? Recent work has shown that children will selectively trust robot teachers: 3–5-year-olds will learn new words from a robot that has previously accurately labeled objects, but not from a robot that has previously inaccurately labeled objects (Brink & Wellman, 2020; Li & Yow, 2023). But it remains an open question as to whether children's social expectations and preferences for human teachers extend to robot teachers. Specifically, are children inclined to trust a robot, even if they are unaware of the robot's competency, and how would children respond to a robot that communicatively signals that a mistake is accidental (or intentional) or that it even feels remorseful for the mistake?

Designing learning systems that mimic human social error-signaling cues (such as marking accidental errors or even apologizing for them) could help mitigate two types of problems: First, it could be beneficial so that learners do not immediately stop trusting technology that occasionally breaks, glitches, or otherwise makes mistakes. But also, socially signaling imperfection could mitigate issues that arise when children (or adults) assume that technologies, like internet search engines, voice assistants, or large language models, are more reliable sources of information than humans by virtue of having easy access to more information (Danovitch & Keil, 2008; Girouard-Hallam & Danovitch, 2022; Robinette, Li, Allen, Howard, & Wagner, 2016; Wang, Tong, & Danovitch, 2019). Either way, understanding the dynamics of trust in learning from social robots across development is an important first step.

To date, questions have primarily been concerned about how much adults trust robots that make errors. The gist of this body of work is that adults distinguish social from technical errors and prefer technologies that act more agent-like. For example, when a robot or chat bot makes technical errors (e.g., typos), adults do not view the technology as human-like and are less likely to maintain trust (Bührke, Brendel, Lichtenberg, Greve, & Mirbabaie, 2021; Westerman, Cross, & Lindmark, 2019). However, when the robot makes social mistakes (e.g., does not follow the rules, makes incongruent gestures, or is overtly mean), adults prefer the robot and think it has human-like qualities (Mirnig et al., 2017; Salem, Eyssel, Rohlfing, Kopp, & Joublin, 2013; Short, Hart, Vu, & Scassellati, 2010; Yasuda, Doheny, Salomons, Sebo, & Scassellati, 2020). Some research suggests that adults also prefer when a robot attempts to rectify its mistakes (e.g., by apologizing or offering compensation; Kim & Song, 2021; Lee, Kiesler, Forlizzi, Srinivasa, & Rybski, 2010; Xu & Howard, 2022).

There are several reasons to think that children will not necessarily respond the same way as adults to robot errors. For one thing, young children are overall more willing to treat robots as agents than adults (Flanagan, Rottman, & Howard, 2021; Jipson & Gelman, 2007; Reinecke, Wilks, & Bloom, 2021). Young children, therefore, might similarly respond to a robot's error as they would for a person's error. Furthermore, children's belief in a technology's agentic capabilities declines with age (Flanagan et al., 2023), so we may even see age-related changes in how children respond to a robot's error. We may also see changes in learning that correspond to changes in learning *goals* across the critical transition from play-based learning to formal, classroom learning. In our formal education system in the U.S., 6-to 7-years-old (first grade) are expected to master foundational skills necessary for learning to read, write, and understand symbolic and numerical systems (Coker & Ritchey, 2010; Rosenkrantz, 2021). Educational technologies, therefore, could be treated differently for children at this age compared to younger children who use technologies more for entertainment purposes. Taking this work together, it is reasonable to anticipate age differences in trust of technologies, both between children and adults and between children of different ages.

In this project, we explored 4–7-year-old children's (Study 1) and adults' (Study 2) trust in technologies during an educational, collaborative game. In the following studies, children and adults played an online, word-learning game with either a human or robot partner. The

game involved 8 Trials in which, for each trial, participants had to guess the "correct" label of a novel object. Critically, before the participants made their own guess, they first heard their partner say what they think is the correct label. We measured trust, therefore, by whether participants picked the label their partner suggested. Importantly, participants got immediate feedback on their and their partner's choice, either indicating that the chosen label was right or wrong. The feedback after each trial, therefore, would theoretically inform participants whether they should trust the human or robot's suggestion on the next trial.

We were first interested in children and adults' baseline trust in their technological partner. Prior work has explored whether children are willing to trust a robot after they see the robot accurately or inaccurately label familiar objects (Brink & Wellman, 2020; Li & Yow, 2023), but it is also important to explore whether children and adults will default to trusting an unfamiliar robot, as they do with other people (Jaswal et al., 2010). Considering this, we intentionally did not give participants any information about their partner's word knowledge. Therefore, participants' responses on the first trial will demonstrate whether children and adults are inclined to trust an unfamiliar robot partner. We can also then see how trust changes as children and adults gain more positive information about their partner through playing the game. So, for Trials 1–4, the robot or human suggested the correct label – if the participant picked the label given by the partner, the participant got the question right; if the participant picked one of the other two labels, the participant got the question wrong.

Our primary interest was whether children and adults lose trust in their partner once it makes a mistake, and whether this depends on the way the partner responds to the loss. To investigate this, for Trials 5–8, the robot or human suggested the wrong label, so that if the participant picked the label suggested, the participant got the question wrong. Children and adults saw the partner consistently react to the inaccuracy in one of three ways: the *Mistaken* partner expressed self-awareness by responding to the loss with shaking their head and saying, "oops I made a mistake" The *Apologetic* partner responded the same as the Mistaken partner but added an apology ("I am so sorry") after admitting the mistake. The *Uncooperative* partner expressed a deceitful intention by responding with pointing their finger and saying, "haha I told you the wrong one."

After all 8 trials, we followed with an interview probing participants' judgments of their partner as well as the other partner that they did not play with. Prior work by Brink and Wellman (2020) found a moderate relationship between children's trust in a robot and believing that the robot has psychological agency (e.g., ability to think, know good from bad). We aim to explore the relationship of psychological agency and trust in light of mistakes. We also include other judgments of agency (e.g., emotional, social, sensory, and competence, similarity to humans) to investigate whether these also relate to children's trust in mistaken technologies.

1. Study 1

1.1. Methods

The study was approved by the Cornell University Institutional Review Board (Protocol IRB0000557) and the Duke University Institutional Review Board (Protocol IRB20220020). The deidentified data set, analysis code, and preregistration for the study are available on the project's Open Science Framework (OSF) page at https://osf.io/n4d23/ (Flanagan et al., 2024).

1.1.1. Participants

The final sample consisted of 168 4–7-year-old children ($M_{age} = 5.96$, $SD_{age} = 1.09$, 48% females) recruited from two lab databases, one in a small city in the Northeastern United States (64% of children) and the other from a city in the Southeastern United States (36% of children). The preregistration had an initial goal of 156 participants (26 in each condition), but an additional 12 participants were collected to have a better distribution of age and gender in each sample. The initial goal of 156 participants was determined through an a priori power analysis using G*Power (version 3.1.9.6; Faul, Erdfelder, Lang, & Buchner, 2007), which found that 154 participants (25.67 in each condition) would be required to achieve 80% power for detecting a medium effect (0.25) at a significance criterion of $\alpha = 0.05$ for an ANCOVA statistical test (number of groups = 6, numerator degrees of freedom = 2). With our new sample number, we conducted a sensitivity power analysis using G*Power and found that 168 participants with a significance criterion of $\alpha = 0.05$ and power = 0.80 would result in a medium effect size (0.24) for the same type of ANCOVA statistical test.

There were 28 children in each condition. Of those that reported, 69% children were White, 17% were Bi- or Multi-racial, 10% were Asian, 3% were Black/African American, 1% was Hispanic/Latino. The majority of children's primary caregivers held a college degree or above (98%) and had a household income of $50,000 or above (91%). Ten additional children participated but were excluded from the final sample due to internet issues ($N = 3$), the child wanting to stop playing the game ($N = 5$), or the child being distracted (e.g., looking away from the computer, talking over videos or the experimenter) throughout the majority of the game ($N = 2$).

1.1.2. Materials

The videos of the robot partner involved a Nao humanoid robot. The Nao robot is 58 cm tall, can speak, and has legs, arms, a torso, and a head with eyes and a mouth. The videos of the human partner involved a young adult woman with blonde hair in a red shirt. Careful attention was paid to making the robot behaviors and vocalizations similar to those of the human partner in the study. To make the partners seem more agentic, the partner's mannerisms included idle behaviors (e.g., slight arm, hand, or head movements) and other possible behaviors to express emotion (e.g., raising its arm towards its chin in a pensive pose; head facing downwards to express disappointment; a thumbs up to express happiness; and a finger pointing to express deceit). The videos of the partners were pre-recorded, but this detail was not mentioned to participants.

There were 8 novel objects used during the word-guessing game, and each object had 3 novel labels (24 novel labels total). The novel objects and labels were obtained through the Novel Object and Unusual Name (NOUN) Database (Horst & Hout, 2016). The entire study, including the word-guessing game and agency questionnaire was developed on Unity and hosted on GitHub Pages as a Unity WebGL build. The web application consists of a graphical user interface (GUI), which includes embedded pictures and videos, that enables the participant to interact with the game and to answer questions.

1.1.3. Procedure

The study was done online via Zoom. Each child sat with their guardian by the computer and the experimenter displayed her screen. When the experimenter displayed her screen, she got confirmation from the child and the child's guardian that they could both see the full screen. Children were then first given a warm-up questionnaire – children were asked how much they like certain foods – to ensure that children could see the screen, to make children feel comfortable in the online testing environment, and to familiarize children with a scale later used in a questionnaire. Children were randomly assigned to one of the six conditions in a 2 × 3 design (Partner type: Human or Robot; Response type: Mistaken or Apologetic or Uncooperative). The study proceeded in two parts: the word-guessing game, then the questionnaire.

First, children were told that they are going to play a word-guessing game with a partner (Anne or Nao). Children were told that in the game they will see an object, Anne/Nao will tell them what she/it thinks it is called, and then the child can make their own guess. Then participants were introduced to the partner via video: the partner waved, said hello and their name (Anne or Nao), and then said, "let's play the game". This

was done to demonstrate that the robot can move and talk on its own, while also establishing that the partner is aware of the game being played. While the video of the partner was still up on the screen, a button that said "next" appeared on the screen and the experimenter repeated the instructions of the game (e.g., "I am going to hit next and you and Nao are going to play the game, where you will see an object, hear what Nao thinks it is called, and you will make your guess.")

There were 8 Trials in the word-guessing game. For each Trial, the participant saw a novel object, three possible novel labels, and a video of the partner on the screen. The partner always gave the first response by saying what she/it thinks the object is called (e.g., "I think it is a lorp"). Then there was a blue arrow that pointed to the label the partner suggested. The participant was then given a chance to endorse the partners' response or pick a different label (e.g., "What about you? Do you think it is called a blap, a lorp, or a tunk? Anne/Nao thinks it is called a lorp."). The order of the partner's suggested label (left, middle, right) was predetermined by a randomized generator and was fixed across participants (L, M, R, L, M, L, M).

For Trials 1–4, if the participant picked the same label as the partner, a bell-ring noise played, and the screen displayed a green check mark and a green arrow pointed to the label the partner suggested. The partner responded to the outcome with encouragement (lifted hands, saying "yay great job"). The experimenter told the child that they and the partner guessed the right answer (e.g., "That was correct. You and Anne/Nao guessed the lorp and that was correct."). If the participant picked a different label, a buzzer noise played and the screen displayed a red X mark, a green arrow pointed to the label that the partner suggested, and a red arrow pointed to the label that the child picked. The partner responded to the outcome with discouragement (shaking head, saying "oh no"). The experimenter told the child that they guessed the wrong answer, but that the partner guessed the right answer (e.g., "That was incorrect. You guessed the blap and that was incorrect. Anne/Nao guessed the lorp and that was correct.").

Starting after the participant's response on Trial 5, and continuing through Trial 8, the feedback was reversed. If participants picked the same label as the partner, then a buzzer noise played and the screen displayed a red X mark, a red arrow pointed to the label that the partner suggested, and a green arrow pointed to one of the other two labels. The experimenter told the child that they and the partner guessed the wrong answer (e.g., "That was incorrect. You and Anne/Nao guessed the koba and that was incorrect."). If the participant picked a different label (either one of the two labels that the partner did not suggest), then a bell-ringing noise played and the screen displayed a check mark, a red arrow pointed to the label that the partner suggested, and a green arrow pointed to the label that the child picked. The experimenter told the child that they guessed the right answer, but that the partner guessed the wrong answer (e.g., "That was correct. You guessed the bosa and that was correct. Anne/Nao guessed the koba and that was incorrect.").

Critically, the partner's response on Trials 5–8 was to her/its own incorrect answer, rather than to the outcome, and varied by condition. In the Mistaken response condition, the partner shook her/its head and said, "oops I made a mistake." In the Apologetic response condition, the partner responded the same as the Mistaken response condition, but added an apology: the partner shook her/its head and said, "oops I made a mistake. I am so sorry." In the Uncooperative response condition, the partner pointed her/its arm and said, "ha ha I told you the wrong one."

After the word-guessing game, participants were told that they were done with the game and were now going to answer a few questions about what they think about their 'partner' and the game they played (see full questionnaire and coding in Supplemental Materials). The questionnaire consisted of questions regarding the partner's mental capabilities (can think for herself/itself), emotions (has feelings like happy and sad), sensations (can see and hear the things around it), friendliness (can be your friend), epistemic knowledge (knows the answers to a lot of questions) and moral knowledge (knows the difference between good and bad). For each question, the screen displayed a picture of the partner, the question, and three possible answers (not at all, a little bit, or a lot). Question order was randomized. Participants were also asked about the partner's ontological status ("Is Anne/Nao more like a person or a computer?"). The screen displayed a picture of the partner, the question, a picture of a human lady at one end of the screen and a picture of a computer at the other end, with tick marks in between as possible answers (person a lot, person a little bit, in the middle, computer a little bit, computer a lot).

After participants answered the questions for the partner they played with, they were then shown the other partner (participants in the Human Partner type were shown the robot partner, participants in the Robot Partner type were shown the human partner). A short video of the partner played in which the partner waved and said hello and their name. Then participants were asked the same questions about the new partner.

At the end of the study, we investigated participants' reasoning about the partner's inaccuracy. Participants were asked, "Remember in the word-guessing game with Anne/Nao. In the beginning Anne/Nao told you the right answers but then Anne/Nao started telling you the wrong answers. Why do you think Anne/Nao told you the wrong answers?" Participants' responses were recorded.

1.1.4. Coding

In the word-guessing game, we measured participants' endorsement of the partner at each trial. Participants received a score of 1 if they picked the same label as the partner and a score of 0 if they picked a different label.

For the agency questionnaire, participants' responses to the mental, emotional, sensory, friendliness, epistemic knowledge, and moral knowledge questions were measured in a 3-point scale coded as 0 (not at all), 1 (a little bit), and 2 (a lot). Participants' response to the ontological status question was measured in a 5-point scale coded as 0 (computer a lot), 1 (computer a little bit), 2 (in the middle), 3 (person a little bit), 4 (person a lot).

Participants' responses to the open-ended question were categorized into any of the following categories, not mutually exclusive. Intention: reference to the partner's actions as either not intentional (e.g., "it was an accident"), intentionally harmful (e.g., "he tried to trick me"), or intentionally helpful (e.g., "she wants us to think for ourselves"). Mechanical Property: reference to the partner's mechanical properties (e.g., "he's a robot", "it's broken"). Competence: reference to the partner's competence (e.g., "she doesn't know the answer"). Game Difficulty: reference to the game being difficult (e.g., "the game got trickier"). Self-blame: reference to the child's role (e.g., "I hit the wrong bottom"). Adult learning: reference to it being an opportunity for the adult to play and learn (e.g., "so I could figure it out myself"). Wrong answer: restating that the partner gave the wrong answer (e.g., "because it was wrong"). Physiology: reference to the partner's physical state (e.g., "she was tired"). Other: any unrelated answer (e.g., "bad") or the child saying they don't know. Two condition blind coders independently categorized all explanations for each response (agreement $\kappa s \geq 0.264$, $ps < 0.001$) and any discrepancies were resolved through discussion.

1.2. Results

We were interested in whether children endorsed the partner's suggested word choices during each phase of the game and whether this differed by the type of partner (human or robot), the partner's response, or both. In the preregistration, we initially planned to compare children's responses for each Trial, but we decided it would be more beneficial to see how children's responses change by Trial as well. The following analyses, therefore, are exploratory, but we report the preregistered analyses in the Supplemental Materials and we summarize any notable findings in the main manuscript.

The Accuracy Phase included all endorsements prior to the partner's first inaccuracy, so does not include the condition-dependent responses

(Trials 1–5). For the Accuracy Phase, therefore, we ran a mixed-effects model with Endorsement as the dependent variable, with partner type (robot or human), Trial (Trials 1–5), and age (in years) as the independent variables, and participant ID as a random intercept. The Inaccuracy Phase included endorsements after the first inaccuracy and through all remaining trials, and thus included all the trials on which children had received condition-dependent feedback of the partner's response to her/its inaccuracy (Trials 6–8). For the Inaccuracy Phase, therefore, we ran a similar mixed-effects model as the Accuracy Phase but included Response type (Mistaken or Apologetic or Uncooperative) as an independent variable. For both of the models, we only included interactions in the model if they were previously found to be significant.

Fig. 1 shows the rates of endorsement of the partner's word choice across the Accuracy phase from Trials 1 to 5. On Trial 1 – after just a brief introduction and prior to corrective feedback – the majority of children followed both the human and robot partner's suggested word choice (Human: 60.7%, $SD = 0.49$; Robot: 61.9%, $SD = 0.49$; binomial $ps < 0.0001$). The final model for the Accuracy Phase included partner type, Trial, and age as factors – interactions were removed from the final model as they were initially found to be insignificant, $ps \geq 0.080$. We did not find a significant main effect of partner type, children trusted the human and robot at similar rates, $\chi^2(1) = 0.38$, $p = .541$. We found a main effect of age, $\chi^2(1) = 7.03$, $p = .008$, such that older children were less likely to endorse the partner's word choices than younger children, $OR = 0.74$, 95% CI (0.59, 0.93). Finally, we found a main effect of Trial, $\chi^2(4) = 55.14$, $p < .001$, such that it took two instances of accuracy for children's trust in the partner to significantly increase above the initial level, whether they were human or robot: Trial 3 versus Trial 2, $OR = 2.96$, $p = .001$, 95% CI (1.64, 5.34). Trust did not differ between the other sequential Trials, $ps \geq 0.237$. Trust remained high throughout the first half of the game (67.9% - 100%), binomial $ps < 0.0001$.

Fig. 2 shows the rates of endorsement of the partner's word choice at Trial 5 and during the Inaccuracy phase from Trials 6 to 8. When the partner starts giving the wrong answer, we found that children's trust varied by the number of inaccuracies, the partner and response type, as well as children's age. In general, children were still willing to trust the partner after one instance of inaccuracy (57.1% -67.9%), Trial 6 binomial $ps < 0.004$. The final model for the Inaccuracy Phase included partner type, Response Condition, Trial, and age as factors as well as two three-way interactions between partner type, Response Condition, and age and between Response Condition, Trial, and age – all other possible two- and three-way interactions were removed from the final model as they were initially found to be insignificant, $ps \geq 0.051$.

Running our model, we found a main effect of age, such that younger children trusted the partner's word choice more than older children, $\chi^2(1) = 9.79$, $p = .002$. We also found a main effect of Trial, $\chi^2(2) = 43.44$, $p < .0001$, such that children's endorsements significantly declined from Trial 6 to 7, $OR = 0.17$, $p < .0001$, 95% CI (0.09, 0.36), but not from Trial 7 to Trial 8, $OR = 0.44$, $p = .051$, 95% CI (0.19, 1.00). The percentage of children trusting the partner at Trial 7 (25% - 42.9%) and Trial 8 (17.9% - 25%) were at chance, binomial $ps > 0.069$, or less than chance (for the Human Uncooperative at Trial 7 (7.1%, $SD = 0.26$) and Trial 8 (10.7%, $SD = 0.32$), binomial $ps < 0.014$. We did not find a main effect of partner type, $\chi^2(1) = 1.80$, $p = .180$, or a main effect of Response Condition, $\chi^2(2) = 4.65$, $p = .098$. We also did not find any significant two-way interactions within the three-way interactions included in our model, $ps \geq 0.337$. However, children's trust in partners who admit mistakes diverged with age, as indicated by two three-way interactions found in the model.

First, we found a significant interaction between partner type, Response condition, and age, $\chi^2(2) = 6.53$, $p = .038$ (see Fig. 3). We looked at this interaction in two ways: how children's trust in each condition changed with age and how trust differed between conditions for each age group. For the former, we ran an additional model with the three-way interaction with age as a continuous variable; for the latter, we ran a similar model but with age as a categorical variable (4–5-year-olds and 6–7-year-olds). All models included Trial and participant ID as random variables. Results of the two models can be summarized as follows: We found that older children were less trusting of a robot that makes a mistake than younger children, $OR = 0.46$, $p = .026$, 95% CI (0.27, 0.78). Older children were also less trusting of a human that apologizes than younger children, $OR = 0.50$, $p = .046$, 95% CI (0.30, 0.83). Furthermore, older children were more trusting of an apologetic robot than an apologetic human, $OR = 3.34$, $p = .034$, 95% CI (1.09, 10.20). Together, these findings suggest that by school age, children are more skeptical of unintentional mistakes made by a robot, but apologies mitigate this effect.

Second, we found a significant interaction between Response condition, Trial, and age, $\chi^2(2) = 15.62$, $p = .004$. Running this additional model, we investigated the differences in children's endorsement by age at each Trial for each Response type, averaged over partner types. At Trial 6, older children were less likely to endorse a Mistaken partner's word choices than younger children, $OR = 0.33$, $p = .029$, 95% CI (0.15, 0.69). We did not find a significant difference in age at Trial 6 for the other Response conditions or at the other trials for any Response condition, $ps \geq 0.073$. This demonstrates that the above finding – that older children are less forgiving than younger children of mistakes – takes effect immediately, after just one instance inaccuracy without apology.

Furthermore, supplemental analyses exploring differences at each Trial (see Supplemental Materials) suggest that children across the age range distinguished between intentional and unintentional inaccuracies, preferring to trust partners whose errors were clearly unintentional. Specifically, children in the Uncooperative conditions were less likely to trust the partner at Trial 7 than children in the Mistaken, $OR = 0.29$, $p = .028$, or Apologetic conditions, $OR = 0.30$, $p = .033$ (see Supplemental Materials for full model results). This aligns with children's open-ended judgments, such that children who played the game with the Uncooperative partner were more likely to say that the partner had harmful intentions than children in the Mistaken or Apologetic conditions (see Supplemental Materials).

Results for children's responses to the agency questions are reported in Supplementary Materials, but we give a brief summary of the findings here. In general, children thought that the human partner was more like a person than the robot. In line with prior work (Flanagan et al., 2023), younger children said that the robot was more like a person than older children. Younger children also said that the robot could have feelings and could think for itself more than older children. Unlike prior work (Brink & Wellman, 2020), we did not find an influence of children's belief that the robot had psychological agency (e.g., could think, know the difference between good and bad) on children's overall trust in the robot. However, we did find that children's belief that the robot knows the answers to a lot of questions increased their overall trust in the robot.

1.3. Discussion

Children in the modern world are surrounded by technologies, particularly in their education. It is critical, therefore, to investigate whether children are trusting of technologies and whether trust can be maintained when the technologies are inaccurate. In this study, we found that 4–7-year-old children are inclined to trust their partner, human or robot, when learning new words, even though children had no prior information on their partner's competency (see Jaswal et al., 2010 for a perspective on young children's default bias to trust). Naturally, children's trust increased as they saw more instances of the partner being accurate (Harris & Corriveau, 2011; Koenig & Harris, 2005b; Koenig & Woodward, 2010; Ronfard & Lane, 2018; Sobel & Kushnir, 2013). When the partner was inaccurate, however, children were sensitive to whether the error was intentional or accidental. Notably, both when the partner was accurate or intentionally inaccurate, children's trust did not differ between the human or robot partner, suggesting that children expect robots to be helpful and accurate teachers just like humans. However, when robots made accidental errors, we found that children's trust

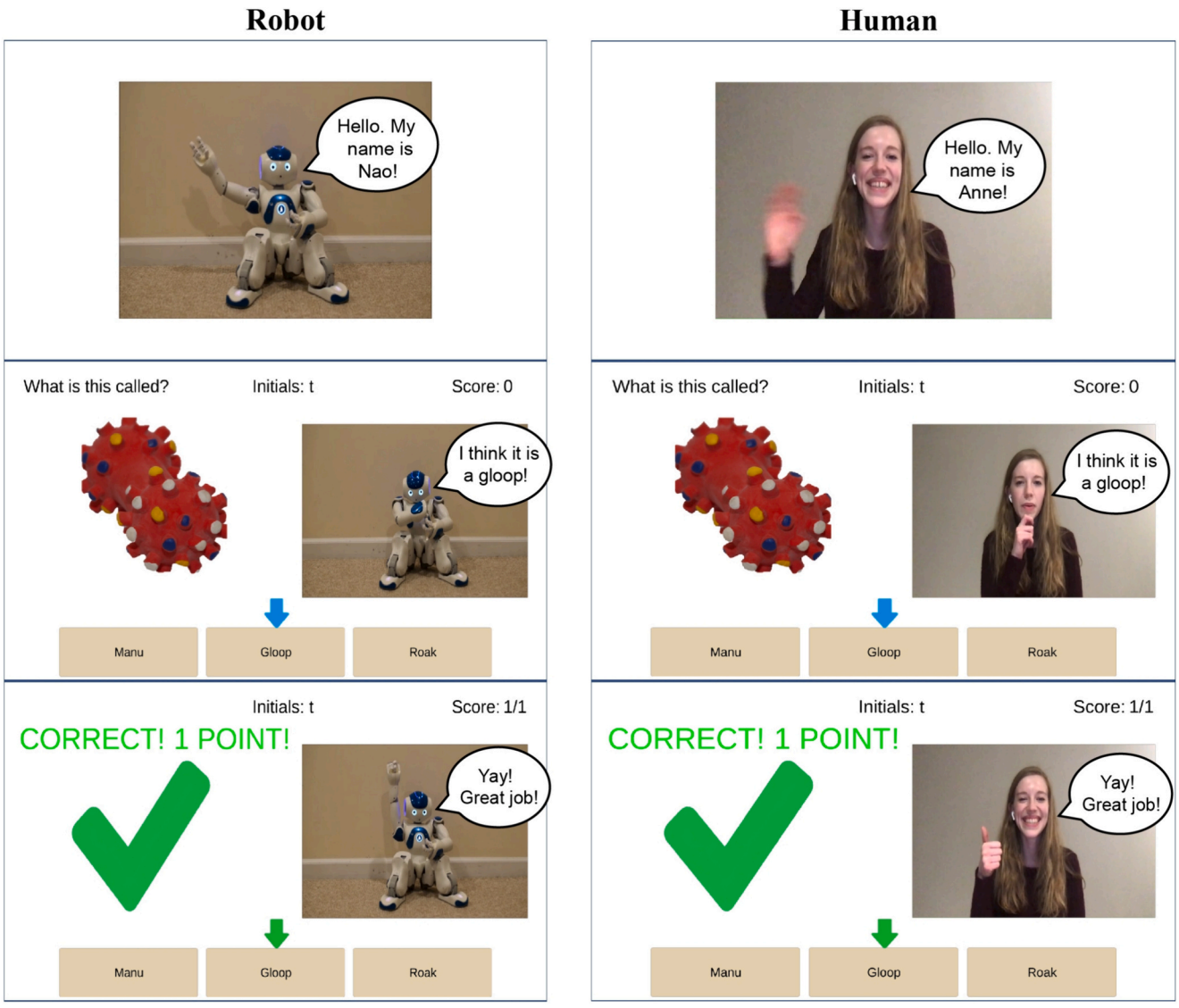

Fig. 1. Example of the word-guessing game during the Accuracy Phase and participant's responses at each Trial (1–5), split by Partner type (Robot and Human) and age (4–5-year-olds, 6–7-year-olds, adults).
Note. Bars represent standard error. For visualization, children's age is grouped categorically, but analyses involve age as a continuous variable.

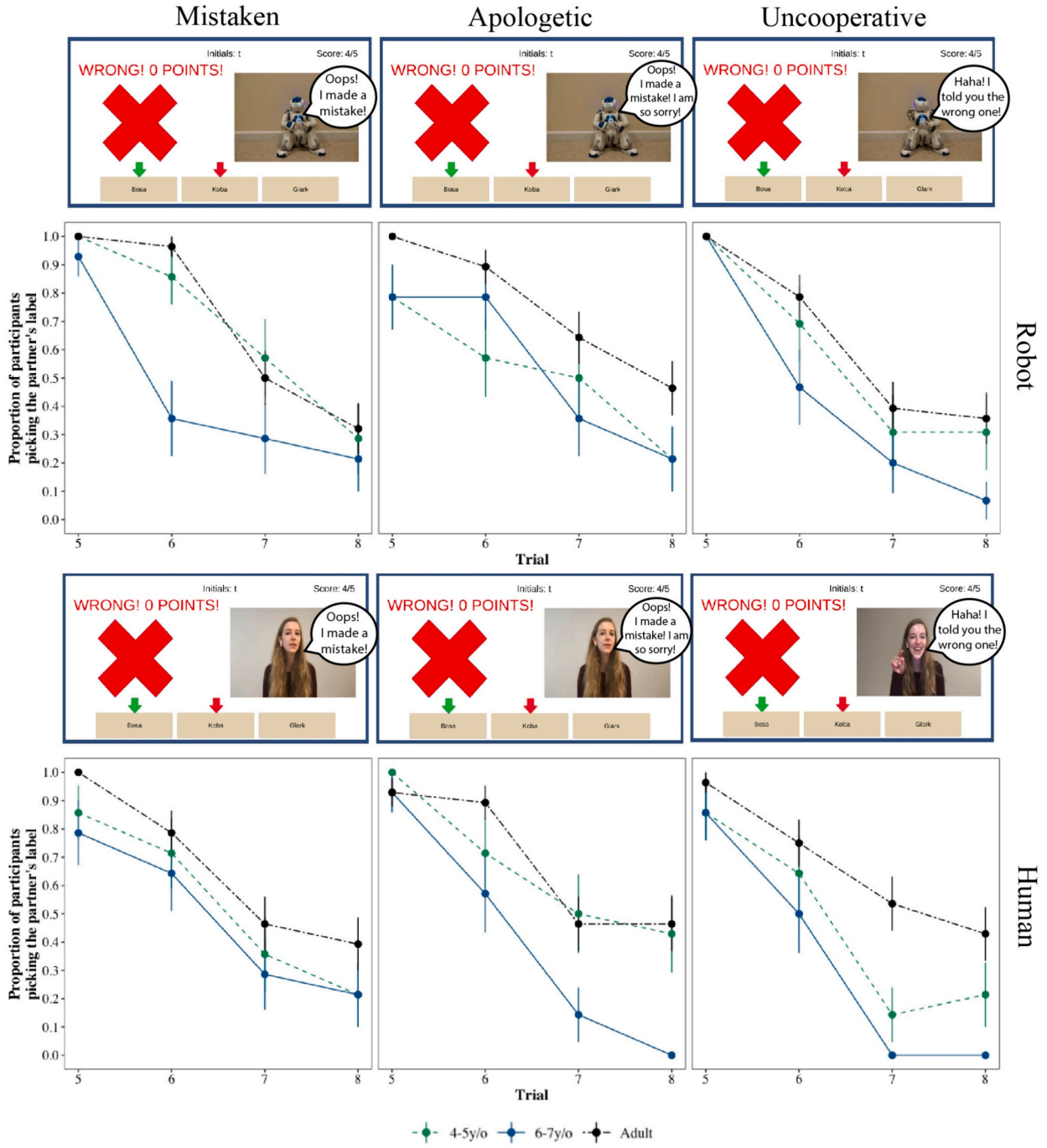

Fig. 2. Example of the word-guessing game during the Inaccuracy Phase and participant's responses at each Trial (5–8), split by Partner type (Robot and Human), Response type (Mistaken, Apologetic, and Uncooperative) and age (4–5-year-olds, 6–7-year-olds, adults).
Note. Bars represent standard error. For visualization, children's age is grouped categorically, but analyses involve age as a continuous variable.

differed with age.

The age-related differences highlight the different ways educational technologies are used in early childhood. In general, we found that older children were less trusting of their partner throughout the entire game, even when the partner was accurate. Considering that children were not given any prior information about their partner before the game, it may be that younger children are inclined to trust unfamiliar partners (Jaswal et al., 2010) while older children may require more information about their partner before deciding to trust. It would be beneficial for work to explore this possibility more directly, as this is relevant to the way children are engaging with educational technologies in their everyday lives: children are often introduced to technologies with little information about the technologies' competence or source of knowledge.

The change in age could also be related to differences in how children engage with educational programs. In preschool, children may use robots and other technology mainly for fun, not for the explicit goal of learning (Schulz & Bonawitz, 2007). As children transition to formal education, they are expected to master foundational skills, such as reading words. As such, children at this age are being evaluated on their academic performance and are increasingly aware, and sensitive to, of their own and others' academic competencies (Bian, Leslie, & Cimpian, 2017; Ruble, Boggiano, Feldman, & Loebl, 1980; Stipek & Daniels, 1988). Older children in our study, therefore, likely took the game more seriously as part of their learning and academic performing. This

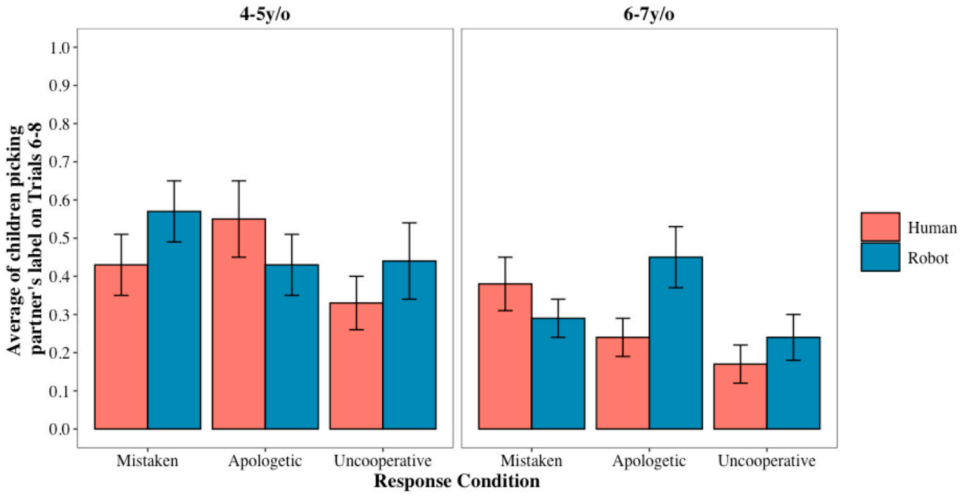

Fig. 3. Mean proportion of children picking the partner's label on Trials 6–8, split by age (4–5-year-olds and 6–7-year-olds), Response type (Mistaken, Apologetic, and Uncooperative) and Partner type (Robot and Human).
Note. Bars represent standard error.

expectation could have made children more restrictive regarding who they trusted.

Furthermore, and most notably, older children were sensitive to the robot's response to the unintentional inaccuracy while younger children seemed to judge the robot's and the human's response similarly. Specifically, older children were more forgiving of a robot that apologizes for its mistake than a human who gives the same apology. An apology from a person can indicate two things: a genuine feeling of remorse or following a social norm. We expect people to apologize after a transgression, so older children may have viewed the human's apology as conforming to norm rather than as sincere (Smith et al., 2018; Waddington et al., 2022). However, we may not expect a machine-like robot to apologize or even be aware of the social norms of apologies. For this reason, older children may have viewed an apology from the robot as a sincere expression of remorse, and perhaps even a promise to not make the same mistake again.

The developmental changes in children's trust of a robot partner call to question how adults would respond. Adults are also using technologies in their daily lives, but are less likely to view technologies as agents than children (Flanagan et al., 2021; Jipson & Gelman, 2007; Reinecke et al., 2021). Furthermore, adults likely use technologies to build upon their existing knowledge (e.g., asking ChatGPT to write an email), rather than using technologies to teach the foundations of their knowledge (e.g., learning new words). For each of these reasons, we may see adults trusting technologies differently than the children in our study. We may also see adults responding differently to the robot's accidental, remorseful, or intentional errors, as following prior work (Kim & Song, 2021; Lee et al., 2010; Xu & Howard, 2022). In the following study, therefore, we investigated adults' trust of a human or robot partner in the same word-learning game as Study 1.

2. Study 2

2.1. Methods

The study was approved by the Duke University Institutional Review Board (Protocol IRB20220350). The deidentified data set, analysis code, and preregistration for the study are available on the project's OSF page at https://osf.io/n4d23/ (Flanagan et al., 2024).

2.1.1. Participants

The final sample consisted of 168 adults ($M_{age} = 33.8$, $SD_{age} = 11.79$, 52% females) recruited from Prolific. There were 28 adults in each condition. Of those that reported, 65% adults were White, 5% were Bi- or Multi-racial, 10% were Asian, 8% were Black/African American, 10% were Hispanic/Latino, 1% was Middle Eastern, and 1% was American Indian. Half of the adults held a college degree or above (57%) and had a household income of $50,000 or above (52%). The same number was determined to match Study 1, but we also conducted a sensitivity power analysis using G*Power and found that 168 participants with a significance criterion of $\alpha = 0.05$ and power = 0.80 would result in a medium effect size (0.24) for an ANCOVA statistical test (number of groups = 6, numerator degrees of freedom = 2).

2.1.2. Materials

The materials used for Study 2 were identical to Study 1.

2.1.3. Procedure

The study was done online in which approved Prolific participants were given a link to the word-guessing game and completed the study unmoderated. Participants were randomly assigned to one of the six conditions in a 2 × 3 design (Partner type: Human or Robot; Response type: Mistaken or Apologetic or Uncooperative). The study proceeded in two parts identical to Study 1: the word-guessing game, then the questionnaire. Since adults' participation was unmoderated, additional items and controls were added to the study to ensure that adults were following the study correctly and paying attention. Specifically, adults had to make the game at full screen before they could participate, video and audio checks were given in the beginning of the study (e.g., adults had to describe video and audio clips), instructions and feedback (matching the exact language the experimenter gave in Study 1) were written at the top of the screen, attention check questions were included throughout (e.g., "Click not at all for this question"), and videos would not play if adults were active on a different window or application on their computer.

2.1.4. Coding

Coding for the word-guessing game and questionnaire in Study 2 was identical to Study 1. For the open-ended question, we removed the category referencing the physiology because no adults gave this explanation. We also included a category for references to intentionally neutral (e.g., "he wanted me to stop trusting him") and study design (e.g., "to test if we would still trust the robot"). Two condition blind coders independently categorized all explanations for each response

(agreement κs ≥ 0.334, *ps* < 0.001) and any discrepancies were resolved through discussion.

2.2. Results

We were interested in whether adults endorsed the partner's suggested word choices during each phase of the game and whether this differed by the type of partner (human or robot), the partner's response, or both. In the preregistration, we initially planned to compare adults' responses for each Trial, but we decided it would be more beneficial to see how adults' responses change by Trial as well. The following analyses, therefore, are exploratory, but we report the preregistered analyses in the Supplemental Materials, and we summarize any notable findings in the main manuscript. The following analyses in the main manuscript were the same as Study 1.

Fig. 1 shows the rates of adult's endorsement of the partner's word choice across the Accuracy phase from Trials 1 to 5. Similar to children, initial rates of endorsements were above chance for both human (82.1%, *SD* = 0.39) and robot (79.8%, *SD* = 0.40), binomial *ps* < 0.0001. The final model for the Accuracy Phase included partner type and Trial as factors – the two-way interaction was removed from the final model as it was initially found to be insignificant, $p = .660$. We did not find a main effect of Partner type, $\chi^2(1) = 0.03$, $p = .868$, demonstrating that adults, like children, trusted both accurate partners at high rates. We found a main effect of Trial, $\chi^2(4) = 31.63$, $p < .001$, such that there was a significant increase in endorsing the partners' label after one instance of accuracy: Trial 2 versus Trial 1, *OR* = 2.50, $p = .047$, 95% CI (1.23, 5.10). Trust did not differ between other sequential Trials, *ps* ≥ 0.481 and remained high throughout the first half of the game (86.9% - 100%), binomial *ps* < 0.0001.

Fig. 2 shows the rates of adult's endorsement of the partner's word choice at Trial 5 and across the Inaccuracy phase from Trial 6 to 8. Adults were even willing to maintain trust after one instance of inaccuracy: adults endorsed the partner's label at high rates at Trial 6 (75% - 96.4%), binomial *ps* < 0.0001. The final model for the Inaccuracy Phase included partner type, Response Condition, and Trial as factors – all two- and three-way interactions were removed from the final model as they were initially found to be insignificant, *ps* ≥ 0.335. We found a main effect of Trial, $\chi^2(2) = 60.24$, $p < .001$, such that adults' endorsements significantly declined from Trial 6 to 7, *OR* = 0.12, $p < .001$, 95% CI (0.05, 0.26), but not from Trial 7 to Trial 8, *OR* = 0.60, $p = .142$, 95% CI (0.50, 1.11). However, looking at our chance comparisons, we never saw a majority of adults mistrust the partner: some adults were still trusting the partner at Trial 7 (39.3% - 50%) and Trial 8 (32.1% - 46.4%), binomial *ps* > 0.069, and a majority of adults were trusting the Human Uncooperative (53.6%, *SD* = 0.51) and Robot Apologetic (64.3%, *SD* = 0.49) conditions at Trial 7, binomial *ps* < 0.026.

We did not find a main effect of partner type, $\chi^2(1) = 0.12$, $p = .730$, or Response Condition, $\chi^2(2) = 2.88$, $p = .238$. Unlike Study 1, even when the partner was intentionally inaccurate, adults trust maintained similarly to the unintentional partners at each trial in the Inaccuracy Phase (see Supplemental Materials). This is interesting considering that adults viewed the uncooperative partner as having malicious intent (see Supplemental Materials).

In an exploratory analysis, we were interested in how adults compared to children. To do this, we created a score of change in trust: first, we created a proportion of trust for the Accuracy Phase (Trials 1–5; e.g., 4/5 = 0.80) and a proportion of trust for the Inaccuracy Phase (Trials 6–8; e.g., 1/3 = 0.33). Next, we subtracted the proportion of trust in the Inaccuracy Phase from the Accuracy Phase (e.g., 0.33–0.80 = −0.47). A higher negative score indicates a greater loss of trust. We ran a General Linear Model with the change of trust score as the dependent variable and age group (4–5-year-olds, 6–7-year-olds, and adults), Response Type, and Partner Type as the independent variables. We found a main effect of age group, $F(2,330) = 3.85$, $p = .022$, $\eta_p^2 = 0.02$. Specifically, adults maintained more trust in the partner (*M* = −0.33, *SD* = 0.33) than 6–7-year-olds (*M* = −0.45, *SD* = 0.31), $t(330) = 2.77$, $p = .016$, *d* = 0.31, 95% CI (0.09, 0.52). Preschool-aged children did not differ between either 6–7-year-olds or adults, *ps* ≥ 0.220. This finding further highlights the nuance of educational technologies for early elementary aged children: an age at which they are presumably using educational technologies more than adults, and yet are less trusting of such technologies than adults.

Results for adults' responses to the agency questions are reported in Supplementary Materials, but we give a brief summary of the findings here. Adults thought the human partner had more agentic capabilities than the robot partner and thought the human partner was more human-like than the robot. Looking at adults' judgments of the robot only, we found that adults' belief that the robot had psychological agency (e.g., can think, know good from bad) and that the robot was more human-like *decreased* adults' overall trust in the robot. Furthermore, when comparing adults' responses to children's, we found that adult viewed the robot partner as less human-like than children. This suggests that even though adults did not view the robot as an agentic being, they still maintained trust in it as they do for humans.

2.3. Discussion

In this study, we investigated whether adults would trust technologies in a word-learning game and whether trust can be maintained once the technologies provide inaccurate information. In general, we found that adults were quick to trust their partner, and this is maintained even when the partner shows that it can be inaccurate. Furthermore, adults' trust did not depend on the partner type or whether the partner's inaccuracy was intentional or accidental. Finally, adults trusted the robot despite reporting that they did not view it has having as much agency as children.

These findings are interesting, considering the differences we found in Study 1 and even prior work has found that adults are sensitive to the partner type and response in their trust in technologies (Kim & Song, 2021; Lee et al., 2010; Xu & Howard, 2022). In these prior studies, however, adults had to trust technologies in more adult-like settings (e.g., investments, driving, service), leading to questions of comparability with our work. It is worth noting, however, that recent work with adults and children participating in selective trust experiments has found that adults maintain trust in human informants longer than children (Ronfard & Lane, 2019). Furthermore, in serious scenarios involving robots as informants, such as when adults have to follow a robot in an emergency evacuation, studies show that trust is maintained despite errors (Robinette et al., 2016). Taking this research along with our own findings raise serious questions as to under what circumstances will adults overtrust technologies to a fault.

3. General discussion

Children and adults are getting a vast amount of their information from technologies. Particularly in education, we are seeing more and more agentic technologies that are designed to teach and collaborate with young children (Breazeal et al., 2016; Chen, Park, & Breazeal, 2020; Hashimoto, Kobayashi, Polishuk, & Verner, 2013; Movellan et al., 2009). The engagement with technologies, from both children and adults, has also increased exponentially in light of the recent COVID-19 pandemic. Therefore, it is critical to investigate how young children and adults are engaging with these technological agents in a learning environment. In this project, we explored whether 4–7-year-old children and adults would trust either a human or robot partner in a word-guessing game. Critically, halfway through the game, the partner started giving the wrong answers to children, either accidentally, remorsefully, or intentionally, and we measured how this inaccuracy changed children's and adults' trust. In general, we found that children and adults were quick to trust a technological agent and maintained trust in the agent after one instance of inaccuracy. However, children lost trust in a

partner that was intentionally inaccurate, while adults did not seem to distinguish between intentional and unintentional errors. Notably, school-aged children (6–7-year-olds) were overall less trusting than adults and, in comparison to preschoolers (4–5-year-olds), were more sensitive to the robot's remorseful response to the unintentional inaccuracy. This finding in particular highlights the different goals and expectations adults and children have when engaging with technologies.

The basis of trust is formed through our goals and experiences with others. We establish collaborative partnership with others when they share the same goals or intentions as ourselves (Grice, 1989; Sperber et al., 2010; Tomasello, 2019). We also incorporate our experiences with people when deciding who or when to trust in various contexts (Harris & Corriveau, 2011; Sobel & Kushnir, 2013). Notably, these goals and experiences can change across development. In our study, we uncovered a U-shaped pattern in the effects of age on trust in technological agents, such that preschoolers and adults were more trusting of an inaccurate partner than 6–7-year-olds. We take these findings to highlight the influence of adults' and young children's different goals and experiences on their trust in technological agents, as we discuss below.

Commonly, young children's selective trust is viewed as "adult-like", but we found that adults were not selective in their trust in our study. Instead, adults maintained trust in their partner, even when the partner was intentionally inaccurate. Recent work by Ronfard and Lane (2019) similarly found that adults were slower to lose trust in an inaccurate informant compared to 4–7-year-old children. They speculated that adults' gradual distrust compared to children could be because adults have more experiences with others – people are rarely ever repeatedly correct or repeatedly incorrect but are instead inconsistent in their accuracy. This could also be the case for adults' rich experiences with technologies, such that adults in our study did not think that a few wrong answers meant that their technological partner would now only be inaccurate. This work, therefore, highlights the importance of taking a developmental approach in investigating our engagement with technologies, and broadly opens the question as to what it means for our trust to be "adult-like".

The similarities and differences between preschoolers' and elementary-aged children's trust in their partner suggest that young children's experiences and goals with technologies differ between ages. Specifically, preschoolers are likey using technologies more as a source of entertainment than information (Girouard-Hallam et al., 2023; Schulz & Bonawitz, 2007), so their goal may have been more focused on playing, rather than learning. This is likely why the uncooperative response type was the only condition in which there were no changes in age: the uncooperative partner was essentially refusing to play with the child as well as refusing to help the child learn.

Finally, early elementary aged children are increasingly aware of the informative nature of technologies (Girouard-Hallam & Danovitch, 2022; Girouard-Hallam et al., 2023) in addition to being more selective in their trust as they get older (Jaswal et al., 2010; Ma & Ganea, 2010; Mills & Elashi, 2014). Taken together, this is likely why we found that 6–7-year-olds were more sensitive to who the partner was and what the partner said when deciding if they should maintain trust in the partner. For example, an apology from a partner that seems like a machine is an unexpected social cue that older children seem to prefer. Using educational technologies, therefore, is not one size fits all. Instead, we need to be mindful of who these technologies are designed for in order for these technologies to be the most beneficial and engaging for children.

The difference in children's and adults' goals and experiences with educational technologies is also evident in how children and adult viewed their robot partner. For example, children's trust in the robot was tied to their belief that the robot was a knowledgeable agent. This finding calls to question if other features of the robot are important to children's trust, and if this changes with age with respect to their goals (e.g., more emphasis on entertainment capacities for younger children, but more emphasis on teaching capacities for older children). In contrast, adults' trust in the robot was tied to their belief that the robot was a machine-like object. These findings suggest that adults may prefer educational technologies to be like sources of online information that they are already familiar with (e.g., the internet; see Wang et al., 2019), while children may require their technologies to be more like human informants (e.g., demonstrating their knowledge, apologizing for their mistakes).

Our project presents new avenues for research on engagement with educational technologies. For example, there are various types of educational technologies (e.g., eBooks, smart speakers, chatbots) that have different features than the humanoid robot used in our studies. Considering that children and adults are engaging with these different forms of educational technologies, it would be fruitful to explore people's responses to various types of mistaken technologies. Furthermore, the game in our studies only involved word learning, but prior work has demonstrated that children trust technologies differently for different domains (e.g., personal knowledge, moral knowledge, Danovitch & Keil, 2008; Girouard-Hallam & Danovitch, 2022). It is likely, therefore, that children and adults may respond to a robot's mistake differently in other domains.

The word-guessing game in our studies was designed to mirror the type of educational games children are playing with already. However, in our study, children and adults only played the game once for a short amount of time (approximately 10 min), while, in the real world, people are playing with educational technologies over extended periods of time and over multiple days. The more experiences children and adults get in these technological spaces likely influences their trust, as there are more instances of the partner being accurate and inaccurate for various reasons. It is unclear, therefore, how children and adults would respond to a mistaken robot if they had more experience with the robot, via a familiarization phase used in prior work (e.g., Koenig & Harris, 2005a) or a longitudinal design.

The word guessing game was also entirely online, to best reflect the types of learning activities children were engaging with during the COVID-19 pandemic. However, the online format does present some limitations. For example, it is unclear if children and adults thought the partner was interacting with them in real-time or in a predetermined way, which might have influenced how they viewed the partner's response to inaccuracy. Furthermore, while we took steps to ensure that adults were paying attention and following the game in Study 2 without an experimenter present, the unmoderated method does leave open questions regarding adults' level of engagement compared to children in Study 1. It also remains an open question if children's and adults' responses in an online study would differ in an in-person study environment, as the current literature has found that children's responses are consistent across study environments for some tasks (Schidelko, Schünemann, Rakoczy, & Proft, 2021), but not others (Lapidow, Tandon, Goddu, & Walker, 2021).

Finally, it is unclear whether our findings are generalizable to the greater global population. The majority of children and adults who participated in the study were White and were in high-income, educated households. Furthermore, all of the participants in this study had access to computers in their home. The children in the study were also comfortable using Zoom. Since the children and adults were already familiar with using technologies, they may have been quicker to trust educational technologies than people who have had little or no experience with technologies. We cannot assume, therefore, that engagement with technologies will be the same for everyone. Instead, research should explore how different individual factors (e.g., race, education, technology experience) play a role in our trust of technologies.

We will continue to get our information about the world from technologies. For children, this will particularly be part of their education: whether technologies are used as an additional learning activity, as part of a school's curriculum, or even as the primary teacher. In order to have these technologies benefit children's education and development, we must investigate all the scenarios in which children could engage with technologies, even when the technologies make mistakes. Our

results in particular uncover the ways in which young children are engaging with educational technologies in the real world. For example, we found that children of all ages are aware of an uncooperative agent's harmful intentions and, thus, do not maintain trust in uncooperative agents. The technologies in children's daily lives, however, do not typically have harmful intentions. Instead, these mistakes are accidental, but how children respond to these accidents depend on their age. Specifically, we found that preschoolers, compared to older children, are more trusting of technological agents, even when the agents are unintentionally inaccurate. For children in formal education, however, these educational games are taken more seriously, and so older children may be incorporating their different expectations of agents (either human or robot) and mistakes (whether an apology is given or not) when deciding whether to maintain trust in an inaccurate technology. Together, this research presents important implications for technology design and education in young children's development.

CRediT authorship contribution statement

Teresa Flanagan: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nicholas C. Georgiou:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Brian Scassellati:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Tamar Kushnir:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Data availability

Deidentified data and analysis code for each study can be found on the project's OSF page at https://osf.io/n4d23/.

Acknowledgments

We thank Nihan Ercanli for her work as an undergraduate research assistant. We also thank Maggie Pan, Ana DeCesare, Sarah Yoon, and Neleh Hopper for their assistants in recruitment and coding. We thank Stephen Parry for providing extensive feedback on data analysis and Lauren Stricker for acting in our videos. We also thank all the families from our community and the adults from Prolific for participating in this project. This research was supported by the National Science Foundation of Learning and Augmented Intelligence (SL) program No. SL-1955280. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at [https://doi.org/10.1016/j.cognition.2024.105814].

References

Amir, D., Ahl, R. E., Parsons, W. S., & McAuliffe, K. (2021). Children are more forgiving of accidental harms across development. *Journal of Experimental Child Psychology, 205*, Article 105081. https://doi.org/10.1016/j.jecp.2020.105081

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist, 48*(5), 505–522. https://doi.org/10.1177/0002764204271506

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*(2), 128–154. https://doi.org/10.1016/j.dr.2010.03.001

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41*(2), 328–337. https://doi.org/10.1037/0012-1649.41.2.328

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics, 3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Bernstein, D., & Crowley, K. (2008). Searching for signs of intelligent life: An investigation of young Children's beliefs about robot intelligence. *Journal of the Learning Sciences, 17*(2), 225–247. https://doi.org/10.1080/10508400801986116

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science (New York, N.Y.)* (Vol. 355, (6323), 389–391. https://doi.org/10.1126/science.aah6524

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

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., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny Valley feelings are acquired in childhood. *Child Development, 90*(4), 1202–1214. https://doi.org/10.1111/cdev.12999

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

Bührke, J., Brendel, A. B., Lichtenberg, S., Greve, M., & Mirbabaie, M. (2021). Is making mistakes human? On the perception of typing errors in chatbot communication. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 4456–4465). http://hdl.handle.net/10125/71158.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science, 8*(1), F13–F20. https://doi.org/10.1111/j.1467-7687.2004.00385.x

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education, 150*, Article 103836. https://doi.org/10.1016/j.compedu.2020.103836

Chernyak, N., & Gary, H. E. (2016). Children's cognitive and behavioral reactions to an autonomous versus controlled social robot dog. *Early Education and Development, 9289*(April), 1–15. https://doi.org/10.1080/10409289.2016.1158611

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics, 113*(4), 708–713. https://doi.org/10.1542/peds.113.4.708

Coker, D. L., & Ritchey, K. D. (2010). Curriculum-based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional Children, 76*(2), 175–193. https://doi.org/10.1177/001440291007600203

Danovitch, J. H., & Keil, F. C. (2008). Young Humeans: The role of emotions in children's evaluation of moral reasoning abilities. *Developmental Science, 11*(1), 33–39. https://doi.org/10.1111/j.1467-7687.2007.00657.x

D'Esterre, A. P., Rizzo, M. T., & Killen, M. (2019). Unintentional and intentional falsehoods: The role of morally relevant theory of mind. *Journal of Experimental Child Psychology, 177*, 53–69. https://doi.org/10.1016/j.jecp.2018.07.013

Doebel, S., & Koenig, M. A. (2013). Children's use of moral behavior in selective trust: Discrimination versus learning. *Developmental Psychology, 49*(3), 462–469. https://doi.org/10.1037/a0031595

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

Fiala, B., Arico, A., & Nichols, S. (2014). You, Robot. *Current Controversies in Experimental Philosophy*, 31–47.

Flanagan, T. M., Georgiou, N., Scassellati, B., & Kushnir, T. (2024, May 14). Children's and adults' trust of inaccurate technologies. Retrieved from osf.io/n4d23.

Flanagan, T., Rottman, J., & Howard, L. H. (2021). Constrained choice: Children's and Adults' attribution of choice to a humanoid robot. *Cognitive Science, 45*(10). https://doi.org/10.1111/cogs.13043

Flanagan, T., Wong, G., & Kushnir, T. (2023). The minds of machines: Children's beliefs about the experiences, thoughts, and morals of familiar interactive technologies. *Developmental Psychology*. https://doi.org/10.1037/dev0001524

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the 3rd international conference on human robot interaction - HRI '08* (p. 145). https://doi.org/10.1145/1349822.1349842

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755. https://doi.org/10.1038/415755a

Girouard-Hallam, L. N., & Danovitch, J. H. (2022). Children's trust in and learning from voice assistants. *Developmental Psychology, 58*(4), 646–661. https://doi.org/10.1037/dev0001318

Girouard-Hallam, L. N., Tong, Y., Wang, F., & Danovitch, J. H. (2023). What can the internet do?: Chinese and American children's attitudes and beliefs about the internet. *Cognitive Development, 66*, 101338. https://doi.org/10.1016/j.cogdev.2023.101338

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences, 25*(10), 896–910. https://doi.org/10.1016/j.tics.2021.07.008

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 366*(1567), 1179–1187. https://doi.org/10.1098/rstb.2010.0321

Hashimoto, T., Kobayashi, H., Polishuk, A., & Verner, I. (2013). Elementary science lesson delivered by robot. In *ACM/IEEE International Conference on Human-Robot Interaction* (pp. 133–134). https://doi.org/10.1109/HRI.2013.6483537

Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust Bias to trust testimony. *Psychological Science, 21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development, 78*(6), 1675–1688. https://doi.org/10.1111/j.1467-8624.2007.01095.x

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., … Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology, 48*(2), 303–314. https://doi.org/10.1037/a0027033

Kahn, P. H., Severson, R. L., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., … Freier, N. G. (2012). Do people hold a humanoid robot morally accountable for the harm it causes?. In *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction - HRI '12* (p. 33). https://doi.org/10.1145/2157689.2157696

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, T., & Song, H. (2021). How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. *Telematics and Informatics, 61*, Article 101595. https://doi.org/10.1016/j.tele.2021.101595

Koenig, M. A., & Harris, P. L. (2005a). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Harris, P. L. (2005b). The role of social cognition in early trust. *Trends in Cognitive Sciences, 9*(10), 457–459. https://doi.org/10.1016/j.tics.2005.08.006

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology, 46*(4), 815–826. https://doi.org/10.1037/a0019664

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction, 13*, 1–9. https://doi.org/10.1016/j.ijcci.2017.04.001

Kushnir, T., & Koenig, M. A. (2017). What i don't know won't hurt you: The relation between professed ignorance and later knowledge claims. *Developmental Psychology, 53*(5), 826–835. https://doi.org/10.1037/dev0000294

Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?" the distinction between causal knowledge and word knowledge guides preschoolers' selective requests for information. *Developmental Psychology, 49*(3), 446–453. https://doi.org/10.1037/a0031649

Landrum, A. R., Eaves, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences, 19*(3), 109–111. https://doi.org/10.1016/j.tics.2014.12.007

Lapidow, E., Tandon, T., Goddu, M., & Walker, C. M. (2021). A tale of three platforms: Investigating Preschoolers' second-order inferences using in-person, zoom, and Lookit methodologies. *Frontiers in Psychology, 12*, Article 731404. https://doi.org/10.3389/fpsyg.2021.731404

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 203–210). https://doi.org/10.1109/HRI.2010.5453195

Li, X., & Yow, W. Q. (2023). Younger, not older, children trust an inaccurate human informant more than an inaccurate robot informant. *Child Development*. https://doi.org/10.1111/cdev.14048. cdev.14048.

Ma, L., & Ganea, P. A. (2010). Dealing with conflicting information: Young children's reliance on what they see versus what they are told. *Developmental Science, 13*(1), 151–160. https://doi.org/10.1111/j.1467-7687.2009.00878.x

McElroy, C. E., Kelsey, C. M., Oostenbroek, J., & Vaish, A. (2023). Beyond accidents: Young children's forgiveness of third-party intentional transgressors. *Journal of Experimental Child Psychology, 228*, Article 105607. https://doi.org/10.1016/j.jecp.2022.105607

Mills, C. M., & Elashi, F. B. (2014). Children's skepticism: Developmental and individual differences in children's ability to detect and explain distorted claims. *Journal of Experimental Child Psychology, 124*, 1–17. https://doi.org/10.1016/j.jecp.2014.01.015

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI, 4*, 21. https://doi.org/10.3389/frobt.2017.00021

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on human robot interaction - HRI '09* (p. 307). https://doi.org/10.1145/1514095.1514189

Myers, L. J., Crawford, E., Murphy, C., Aka-Ezoua, E., & Felix, C. (2018). Eyes in the room trump eyes on the screen: Effects of a responsive co-viewer on toddlers' responses to and learning from video chat. *Journal of Children and Media, 12*(3), 275–294. https://doi.org/10.1080/17482798.2018.1425889

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*(5), 722–731. https://doi.org/10.1111/j.1467-7687.2008.00722.x

Oostenbroek, J., & Vaish, A. (2019). The emergence of forgiveness in young children. *Child Development, 90*(6), 1969–1986. https://doi.org/10.1111/cdev.13069

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

Reinecke, M. G., Wilks, M., & Bloom, P. (2021). Developmental changes in perceived moral standing of robots. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (p. 43). Retrieved from https://escholarship.org/uc/item/8f32d068.

Rizzo, M. T., Li, L., Burkholder, A. R., & Killen, M. (2019). Lying, negligence, or lack of knowledge? Children's intention-based moral reasoning about resource claims. *Developmental Psychology, 55*(2), 274–285. https://doi.org/10.1037/dev0000635

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 101–108). https://doi.org/10.1109/HRI.2016.7451740

Ronfard, S., & Lane, J. D. (2018). Preschoolers continually adjust their epistemic trust based on an Informant's ongoing accuracy. *Child Development, 89*(2), 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*, Article 104662. https://doi.org/10.1016/j.jecp.2019.104662

Rosenkrantz, H. (2021). *What should a first grader know?* U.S. News. October 5 https://www.usnews.com/education/k12/articles/what-should-a-first-grader-know.

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology, 16*(2), 105–115. https://doi.org/10.1037/0012-1649.16.2.105

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(−like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics, 5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Schidelko, L. P., Schünemann, B., Rakoczy, H., & Proft, M. (2021). Online testing yields the same results as lab testing: A validation study with the false belief task. *Frontiers in Psychology, 12*, Article 703238. https://doi.org/10.3389/fpsyg.2021.703238

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! An interaction with a cheating robot. In *5th ACM/IEEE international conference on human-robot interaction, HRI 2010* (pp. 219–226). https://doi.org/10.1145/1734454.1734546

Smith, C. E., Anderson, D., & Straussberger, A. (2018). Say You're sorry: Children distinguish between willingly given and coerced expressions of remorse. *Merrill-Palmer Quarterly, 64*(2), 275. https://doi.org/10.13110/merrpalmquar1982.64.2.0275

Smith, C. E., Chen, D., & Harris, P. L. (2010). When the happy victimizer says sorry: Children's understanding of apology and emotion. *British Journal of Developmental Psychology, 28*(4), 727–746. https://doi.org/10.1348/026151009X475343

Smith, C. E., & Harris, P. L. (2012). He Didn't want me to feel sad: Children's reactions to disappointment and apology. *Social Development, 21*(2), 215–228. https://doi.org/10.1111/j.1467-9507.2011.00606.x

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120*(4), 779–797. https://doi.org/10.1037/a0034191

Sommer, K., Davidson, R., Armitage, K. L., Slaughter, V., Wiles, J., & Nielsen, M. (2020). Preschool children overimitate robots, but do so less than they overimitate humans. *Journal of Experimental Child Psychology, 191*, 104702. https://doi.org/10.1016/j.jecp.2019.104702

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stipek, D. J., & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment? *Journal of Educational Psychology, 80*(3), 352–356. https://doi.org/10.1037/0022-0663.80.3.352

Strouse, G. A., & Ganea, P. A. (2017). Toddlers' word learning and transfer from electronic and print books. *Journal of Experimental Child Psychology, 156*, 129–142. https://doi.org/10.1016/j.jecp.2016.12.001

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A Meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development, 92*(1). https://doi.org/10.1111/cdev.13429

Strouse, G. A., Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology, 166*, 310–326. https://doi.org/10.1016/j.jecp.2017.09.005

Tomasello, M. (2019). *Becoming human*. Cambridge, MA: Harvard University Press.

Troseth, G. L. (2003). Getting a clear picture: Young children's understanding of a televised image. *Developmental Science, 6*(3), 247–253. https://doi.org/10.1111/1467-7687.00280

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young Children's understanding of video. *Child Development, 69*(4), 950–965. https://doi.org/10.1111/j.1467-8624.1998.tb06153.x

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*(1), 124–130. https://doi.org/10.1348/026151010X532888

Waddington, O., Jensen, K., & Köymen, B. (2022). Boundaries of apologies: Children avoid transgressors who give the same apology for a repeat offence. *Cognitive Development, 64*, Article 101264. https://doi.org/10.1016/j.cogdev.2022.101264

Wang, F., Tong, Y., & Danovitch, J. (2019). Who do I believe? Children's epistemic trust in internet, teacher, and peer informants. *Cognitive Development, 50*, 248–260. https://doi.org/10.1016/j.cogdev.2019.05.006

Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I Believe in a Thing Called Bot: Perceptions of the Humanness of "Chatbots". *Communication Studies, 70*(3), 295–312. https://doi.org/10.1080/10510974.2018.1557233

Xu, J., & Howard, A. (2022). Evaluating the impact of emotional apology on human-robot trust. In *2022 31st IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 1655–1661). https://doi.org/10.1109/RO-MAN53752.2022.9900518

Yasuda, S., Doheny, D., Salomons, N., Sebo, S. S., & Scassellati, B. (2020). Perceived Agency of a Social Norm Violating Robot. In *Proceedings of the annual meeting of the cognitive science society*. Retrieved from https://par.nsf.gov/servlets/purl/10284325.