Int. J. Human–Computer Studies 177 (2023) 103063

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

Transparent robots: How children perceive and relate to a social robot that acknowledges its lack of human psychological capacities and machine status

Caroline L. van Straten *, Jochen Peter, Rinaldo Kühne

Amsterdam School of Communication Research (ASCoR), University of Amsterdam, P.O. box 15791, 1001 NG Amsterdam, the Netherlands

A R T I C L E I N F O

Keywords:
Child-robot interaction
Child-robot relationship formation
Transparency
Social robotics
Responsible robotics

A B S T R A C T

Children will increasingly encounter, and form social relationships with, social robots. Accordingly, scholars have called for transparency toward children about what social robots are and what they can(not) do to manage children's expectations of this new type of communication partner. Prior research has shown that the way adults present social robots to children can influence children's perception of, and relationship formation with, a robot. To date, however, no studies have yet investigated whether a social robot's own provision of transparent information about its (in)abilities can alter how children perceive and relate to it. To fill this gap initially, we conducted a one-factorial between-subject experiment among 276 children aged 8–10 years old. Children interacted with a robot that either provided them with information about its lack of human psychological capacities and machine status, or not. Exposure to this information decreased children's feelings of closeness toward and trust in the robot. Children's tendency to anthropomorphize the robot mediated the effects of transparency on closeness and trust, while their perception of the robot's similarity to themselves only mediated children's feelings of closeness. Our findings are discussed in light of the ongoing ethical discussion on child-robot relationships.

1. Introduction

The chances of children encountering social robots, or robots *perceived* as being social (van Wynsberghe, 2021), in their everyday lives are gradually increasing (Pashevich, 2021). Accordingly, both the number of studies on the benefits of child-robot interaction (CRI) applications and the number of studies that critically evaluate the desirability of such applications are quickly growing (Pashevich, 2021; see also Prescott and Robillard, 2021). Concerns center, among other things, around the emergence of social relationships between children and robots and the potential consequences of such relationships on children's socio-emotional development (e.g., Borenstein and Pearson, 2013; Turkle, 2006). At the core of these concerns lies the notion that social robots and the relationships they evoke are intrinsically deceptive, or inauthentic, as they (un)intentionally create the illusion of robots being more capable and, in fact, more social than they actually are (e.g., Sharkey and Sharkey, 2020; Turkle, 2007).

Scholars tend to be particularly concerned about children. While the deception of adults "could be accounted [for] by providing an explanation of what [a] robot is capable of and how [its] technology is constructed, in children, critical technological thinking is difficult to induce" (Tolksdorf et al., 2020, p. 134). Still, several studies have indicated that children's perceptions of, and relationship formation with, social robots can at least partially be influenced by how social robots are presented to children (e.g., Cameron et al., 2017; Chernyak and Gary, 2016; Somanader et al., 2011; Tozadore et al., 2017; Van Straten et al., 2020c, 2022). In these studies, adults explained or demonstrated robots' machine status and limited (autonomous) capacities to children before, during, or after their interaction with a robot. To date, however, it has not yet been investigated whether such explanations could also effectively be provided by a social robot itself. As the robots that children encounter will, in the future, increasingly function autonomously (e.g., Stapels and Eyssel, 2021), it seems timely to gain insight into the efficacy of this approach.

Against this background, the present study investigates whether information about a social robot's machine status and lack of human

* Corresponding author at: Amsterdam School of Communication Research (ASCoR), University of Amsterdam, P.O. box 15791, 1001 NG Amsterdam, The Netherlands.

E-mail addresses: c.l.vanstraten@uva.nl (C.L. van Straten), j.peter@uva.nl (J. Peter), r.j.kuhne@uva.nl (R. Kühne).

https://doi.org/10.1016/j.ijhcs.2023.103063
Received 20 December 2022; Received in revised form 26 April 2023; Accepted 29 April 2023
Available online 29 April 2023
1071-5819/

psychological capacities can alter how children perceive and relate to a robot if such transparent information is provided by the robot itself, thereby extending the research by Van Straten et al. (2020c). In doing so, the study contributes to the emerging research field of robot transparency (for a review, see Schött et al., 2023) and responds to calls for responsible robotics that emphasize the importance of openness about social robots, as well as of the anticipation of potential adverse consequences of human-robot interaction (HRI; e.g., the principles for responsible robotics by Boden et al., 2017; the European Commissions Artificial Intelligence Act, Schaake, 2021; the Montreal Declaration for responsible AI development, Université de Montréal, 2018). More specifically, the study investigates the effectiveness of transparency as a means to purposefully alter children's responses to social robots (as suggested, e.g., by Scheutz, 2012).

2. Theoretical framework

2.1. Effects of transparency on children's responses to social robots

In calls for responsible robotics, scholars have argued that robots' machine nature and working should be transparent to their users (see, e.g., Boden et al., 2017). The goal of such an approach is to decrease users' deception about the features and abilities of robots. Several CRI studies can be interpreted as having investigated the effects of transparency. For example, various researchers have explained to children that robots were being teleoperated (i.e., manually controlled) during an interaction, thus revealing the so-called Wizard-of-Oz (WOZ) technique. Their results consistently showed that transparency about a robot's teleoperation can decrease young children's (i.e., aged 7 and under) perception of robots' humanlikeness (Cameron et al., 2017), sentience, moral standing (Chernyak and Gary, 2016), and their possession of memory and vision (Somanader et al., 2011).

In contrast, studies with children older than 7 initially found no effects of demonstrating social robots' machine nature and working, during or after CRI, on children's relationship formation with, and perceptions of, social robots (i.e., Bumby and Dautenhahn, 1999; Cameron et al., 2017; Turkle et al., 2006). Yet, two more recent studies among children aged 7–10 reported some influences of revealing a robot's teleoperated working on these children's robot perceptions too: First, children perceived a robot as less intelligent after they were told that it had been remotely controlled during their interaction with it (Tozadore et al., 2017). Second, and in line with this finding, children rated a robot lower in autonomy and anthropomorphized it less when they were told *upfront* that the robot would be manually controlled during an interaction (Van Straten et al., 2022).

In another study, an adult provided children, during their interaction with a social robot, with information about the robot's lack of human psychological capacities. Such information led children to anthropomorphize the robot less and decreased children's trust in the robot as well as their perception of its animacy, social presence, and similarity to themselves (Van Straten et al., 2020c). Together, these findings show that transparency about a robot's machine status and working may influence, at least partially, children's robot perceptions as well as child-robot relationship formation.

2.2. Robots as the source of transparent information

Children's interactions with social robots may not always be supervised by adults. Prior research, however, has primarily addressed whether and how the manner in which adults present robots to children affects children's perceptions of, and relationship formation with, a social robot (e.g., Bumby and Dautenhahn, 1999; Kory Westlund et al., 2016; Turkle et al., 2006; Van Straten et al., 2020c, 2022). It is crucial to understand how children's conceptions of social robots can be influenced by transparent information that is offered by adults. At the same time, Schött et al. (2023, Fig. 2, p. 9) outline that robot transparency can not only be conveyed by having others give information about a robot ('transparency on the robot'), but also by the robot itself ('transparency through the robot'). Robots that provide transparent information about themselves mirror (future) child-robot interactions more realistically than adult-led or adult-supervised interactions. In this context, it is important to also investigate whether social robots themselves can effectively communicate transparent information to children.

The question whether a robot's provision of transparent information about itself may be effective is strongly linked to the notion of source credibility. Source (or communicator) credibility refers to "judgements made by a perceiver […] concerning the believability of a communicator" (O'Keefe, 2002, p. 181), and is influenced, among other things, by the communicator's knowledge and authority (Rieh, 2016). Children are likely to consider adults (more than robots) to be authority figures, who have more knowledge than they possess themselves. As a consequence, children may attribute greater credibility to information coming from an adult than to information coming from a robot. In line with this idea, Edwards et al. (2016) found that college students perceived a human teacher as more credible than a robot teacher. Accordingly, students learned better from the human than from the robot (Edwards et al., 2016).

A recent study, however, found that children aged 3 to 6 would consult a Nao robot rather than an adult for answers to machine-related questions (Oranç and Küntay, 2020). This finding dovetails with Vanderborght and Jaswal's (2009) finding that preschoolers consulted children rather than adults for toy-related information. Apparently, children judge the credibility of information sources depending on the type of information they seek. By extension, children may thus consider the robot a reliable source of information about its own machine status and lack of human capacities.

To the best of our knowledge, no studies have yet investigated how children in middle childhood (i.e., 6–12 years old; Cole et al., 2005) conceive of social robots' source credibility. Still, Cameron et al. (2017) found that children aged 6 and younger considered a robot to be person- rather than machinelike, while children aged 6 and older conceived of the robot as a machine. Accordingly, children in middle childhood may consider a robot to be a competent and reliable source of information about its own machine nature. In this developmental period, but more specifically from the age of 8, children gradually develop critical thinking skills and "compare everything in their surroundings against standards of genuineness and authenticity" (Valkenburg and Piotrowski, 2017, p. 70). It thus seems likely that a robot's provision of transparent information about itself will, at least to some extent, affect how children aged 8 and older (who are the target group of the present study, see below) perceive and relate to it.

2.3. Robots' lack of human psychological capacities

Following Van Straten et al.'s (2020c) approach, we investigated the effects of information about a social robot's lack of five human psychological capacities that Hubbard (2011) identified as requirements for a machine to be granted personhood: intelligence, self-consciousness, identity construction, emotionality, and social cognition. As Scheutz (2012, p. 218) outlined, transparency about a robot's lack of such capacities "might reduce the likelihood and extent to which [children] form emotional bonds with robots". In addition, Severson and Carlsson (2010) discussed how children's reasoning about social robots seems to allow for contradictions: Even if they understand that robots lack *biological* characteristics (e.g., animacy), children still attribute *psychological* characteristics to social robots (e.g., intelligence, emotionality). Being transparent about a robot's lack of human psychological capacities, rather than about its lack of biological ones, may thus be most effective when the aim is to alter children's reasoning about, and relationship formation with, social robots.

Van Straten et al. (2020c, p. 6) argued that transparent information about a robot's lack of psychological capacities should not be provided by the robot itself, as “having the robot provide the explanations itself would [...] suggest that the robot has knowledge of its ’conceptual self’ – which implies self-consciousness”. Indeed, having the robot as the source of such information may, to some extent, weaken its effects. However, as mentioned above, CRI under direct and continuous adult supervision may in the future be the exception rather than the norm. Therefore, we consider it – also in line with broader developments in research on robot transparency (Schött et al., 2023) – timely and important to investigate whether a robot’s provision of transparent information about itself may be effective.

Two concepts that play a central role in the emergence of interpersonal relationships are closeness and trust (Bauminger-Zviely and Agam-Ben-Artzi, 2014; Berscheid and Regan, 2005). Closeness can be described as a feeling of connectedness or intimacy that may eventually develop into friendship (Sternberg, 1987), while trust refers to a belief in the benevolence and honesty of another person (Larzelere and Huston, 1980). Van Straten et al. (2020c) found an effect of transparent information about a robot’s lack of human psychological capacities on children’s trust in the robot, but not on their feelings of closeness. The authors interpreted this as indicative of children’s tendency to experience closeness toward objects, media characters, and imaginary others throughout middle childhood (e.g., Gleason, 2013).

An alternative explanation, however, may be that children’s closeness toward the robot remained unaffected because the information they received was offered by the adult rather than the robot itself. Consequently, the positive influence of the robot’s own contributions to the interaction on children’s experience of closeness may have outweighed the possibly negative influence of the transparent information. In the present study, transparent information is offered by the robot itself throughout the interaction and may thus more directly influence children’s closeness toward the robot. While we also expect a decrease in children’s trust in the robot as a function of its provision of transparent information, this does not contradict the idea that children consider the robot to be a credible information source: The belief in another person’s trustworthiness as a friend does, after all, not necessarily coincide with the person’s credibility when providing information regarding a particular topic. Our first hypothesis thus predicts:

Hypothesis 1 (H1). When a social robot informs children about its lack of human psychological capacities and machine status, this decreases children’s feelings of (a) closeness toward and (b) trust in the robot as compared to when the robot does not provide such information.

Based on Van Straten et al.’s (2020c) findings, we moreover expect that transparent information about a robot’s lack of Hubbard’s (2011) human psychological capacities will negatively affect children’s tendency to anthropomorphize the robot and their perception of the robot’s similarity to themselves. Anthropomorphism denotes “the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions” (Epley et al., 2007, p. 864). Anthropomorphization thus comprises a series of characteristic attributions that transparent information about a robot’s lack of human psychological capacities specifically should address. Finally, anthropomorphic tendencies are closely related to the perception of an agent as being similar to oneself (Epley et al., 2007). Perceived similarity is, as a concept, both narrower and broader than anthropomorphism: It is narrower because it involves the comparison of the robot to the self, rather than to humans in general. It is broader because it does not focus on the attribution of humanlike traits but on the perception that the robot is, more generally, ‘like oneself’. Both concepts are central to children’s thinking about robots as more or less humanlike others. Our second hypothesis therefore states:

Hypothesis 2 (H2). When a social robot informs children about its lack of human psychological capacities and machine status, this decreases children’s tendency to anthropomorphize the robot as well as their perception of its similarity to themselves, as compared to when the robot does not provide such information.

2.4. Processes underlying child-robot relationship formation

To date, little is known about the psychological mechanisms that may underlie child-robot relationship formation (Van Straten et al., 2020b). While the development of children’s feelings of closeness toward, and trust in, robots has received growing research attention (for reviews, see Stower et al., 2021; Van Straten et al., 2020b), we do not understand well why and how child-robot relationship formation unfolds. In the broader field of HRI research, however, several studies have found people’s robot perceptions to mediate their feelings of closeness toward and trust in social robots (e.g., Kim et al., 2013; Lee, Jung, et al., 2006, Lee, Peng, et al., 2006). Likewise, it is conceivable that children’s perception of a robot plays a role in the emergence of child-robot relationships. More specifically, there are reasons to believe that children’s perception of a robot as a humanlike entity that is similar to themselves is associated with the emergence of closeness and trust.

The importance of perceived similarity to relationship formation has been well-documented – both regarding interpersonal relationships in general (see Montoya et al., 2008 for a meta-analysis) and the development of friendships in (middle) childhood in particular (see Hartup et al., 2006). Although children’s peer friendships cannot be equated with children’s relationships with social robots, perceived similarity may also be associated with child-robot relationship formation given its importance to the reduction of uncertainty about others: When we perceive someone else to be similar to ourselves, it becomes easier to determine what we can expect from this person – which, according to Uncertainty Reduction Theory, facilitates the emergence of social relationships (see Berger and Calabrese, 1975). As social robots are still relatively new to most children, children are usually quite uncertain as to what can be expected of a robot (Paiva et al., 2018). Therefore, children’s perception of a robot’s similarity to themselves may be positively associated with the initial emergence of child-robot relationships in terms of children’s feelings of closeness toward, and trust in, the robot.

Likewise, children’s tendency to anthropomorphize a robot may be associated with child-robot relationship formation. In a study on children’s (aged 8–14) perception of various kinds of humanoid robots, Tung (2016) found that children’s social attraction toward robots increased with the robots’ more humanlike appearance. As robots’ humanlike appearance is related to children’s tendency to anthropomorphize robots (Manzi et al., 2020), this tendency may thus be associated also with children’s feelings of closeness toward a robot. In addition, a recent meta-analytic review found a positive link between adults’ anthropomorphic reasoning about a social robot and their trust in it (Hancock et al., 2020). While children’s responses to robots may differ from those of adults, it is conceivable that children’s tendency to anthropomorphize a robot may similarly correspond with their trust in it.

Friendships normally develop between more or less equal entities (Emmeche, 2014) that choose to enter (and can also choose to leave) a relationship (Keller, 1997). Both anthropomorphism and perceived similarity tap into children’s reasoning about the robot as an ‘equal other’. Rather than just being *associated*, children’s anthropomorphic tendencies and perceptions of the robot’s similarity to themselves may thus, at least partly, *explain* children’s feelings of closeness toward, and trust in, the robot. In order words, anthropomorphism and perceived similarity may mediate the effects of transparent information on child-robot relationship formation. Therefore, our third and fourth hypotheses predict (see Fig. 1 for a visualization of all hypotheses):

Hypothesis 3 (H3). Children’s tendency to anthropomorphize the robot and their perception of its similarity to themselves will be positively related to their feelings of (a) closeness toward and (b) trust in the robot, such that a decrease in children’s anthropomorphic tendencies and perceived similarity will be associated with a decrease in children’s feelings of closeness and trust.

Hypothesis 4 (H4). Based on H1, H2, and H3, we expect a negative

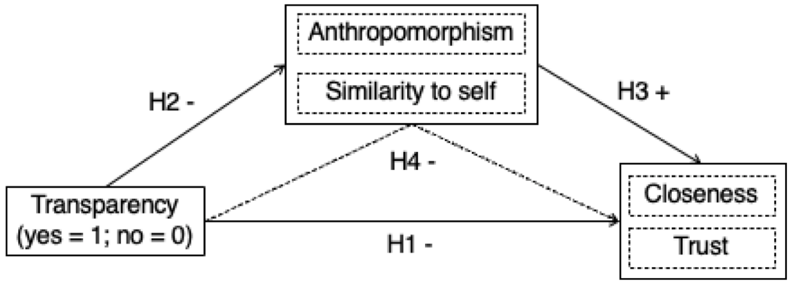

Fig. 1. Conceptual model of hypothesized relationships.

indirect effect: The negative effects of a robot's provision of information about its lack of human psychological capacities and machine status on closeness and trust are transmitted by children's decreased anthropomorphic tendencies and lower ratings of the robot's similarity to themselves.

3. Method

We conducted a preregistered (osf.io/3867k) one-factorial experiment with the robot's sharing of information about its' lack of human psychological capacities and machine status (yes/no) as the between-subject factor. The data were collected in August 2021, at the Research and Development Lab of the NEMO Science Museum Amsterdam.

Before we started collecting the data, ethical approval was obtained from the Ethics Review Board of the Faculty of Social and Behavioral Sciences of the University of Amsterdam. Children's parents were asked to provide written informed consent via a secured online consent form after reading an information letter that explained to them the experimental goals and procedure. We asked parents to indicate, on the consent form, whether their child was diagnosed with autism spectrum disorder (ASD). Individuals who have ASD tend to differ in their anthropomorphic tendencies (see Atherton and Cross, 2018 for a review), and typically experience difficulties with respect to social relationships (see Petrina et al., 2014 for a review). Therefore, children who have ASD may respond differently to social robots than other children. While they could participate in the study, the questionnaire procedure was optional for these children as their data would be excluded from analyses. Parents were informed about this in the information letter.

3.1. Participants

To estimate the required power for testing H1 and H2, we conducted two *a priori* power analyses using G*power (for *F*-tests, each with a power of 0.8 and an alpha level of 0.05; Faul et al., 2007) based on the effect sizes we found (for the variables included in this study) in an earlier experiment with a similar manipulation (see Van Straten et al., 2020c). In terms of H1, the effect sizes found in Van Straten et al. (2020c) suggest a sample size of $N = 93$ is required to identify a main effect on children's trust in the robot as well as an $N = 780$ to find a main effect of the manipulation on children's closeness toward the robot. In terms of H2, a sample size of $N = 7$ would be needed to detect a direct effect on anthropomorphism and $N = 25$ is needed for perceived similarity. For H3, we calculated the required sample size based on Monte Carlo confidence intervals using the R-based application developed by Schoemann et al. (2017). With trust as the dependent variable, we would need a sample size of $N = 1330$ to test the mediating effect of anthropomorphism and a sample size of $N = 255$ to test the mediating effect of perceived similarity. With closeness as the dependent variable, we would need a sample size of $N = 419$ to test the mediating effect of anthropomorphism and a sample size of $N = 46$ to test the mediating effect of perceived similarity.

We had a set agreement on the number of days we could be present at the museum. In the preregistration (osf.io/3867k), we described that, based on estimated visitor numbers, we expected to be able to collect data from 171 children at most. In the end, we were able to collect data from 290 children aged 8 to 10 years old. Although the obtained sample size was thus considerably larger than expected, our study was underpowered to test some of our hypotheses (i.e., the main effect on closeness and some mediation paths). However, as the processes underlying child-robot relationship formation are underexplored, we still considered it informative to analyze the expected relationships. The upper age boundary of 10 years was decided upon to limit developmental differences between children in our sample (e.g., their ability to distinguish what is real from what is not, see Section 2.2). We excluded the data of eleven children who had ASD. In addition, the data of three children were excluded because serious technical problems had occurred with the robot set-up that interfered with a proper execution of the experiment.

We thus analyzed the data of 276 children (142 girls, 134 boys; $M_{age} = 9.50$; $SD_{age} = 0.84$). When children indicated that they did not know how to answer certain questionnaire items, we entered their answers to these items as missing values. In case children did not know how to answer any item of a particular scale, they were thus excluded from the analysis of the respective measure. Participants were randomly assigned to the experimental conditions. There were no significant differences in children's age (t (274) $= -0.148$, $p = .883$) or self-reported gender (χ^2 (1, $N = 276$) $= 0.017$, $p = .896$) across the conditions. Thus, the randomization procedure was successful.

3.2. Interaction task & manipulation

Each child engaged in one 6- to 7-minute, one-on-one interaction with the Nao robot (Softbank; see Fig. 2). The interaction was intentionally kept short, to ensure that participation in the experiment would minimally interfere with children's museum visit. The interaction begun with a short introduction phase during which the robot and child introduced themselves. Then, they played a guessing game during which the robot made seven assertions (e.g., "My favorite color is red") of which the child had to guess whether they were true or false (similar to the game used in Van Straten et al., 2020a). To make the game less repetitive, the robot also told children a riddle. During the interaction, the robot stood entirely still without blinking to prevent anything other than the interaction content from influencing children's robot perceptions. At the end of the interaction, the robot said that the child had done a good job at the game, wished him/her a good time at the museum, and sat down (i.e., was put in standby mode).

Each time the child had guessed whether an assertion was true or false, as well as following the riddle, the robot provided the child with some explanation regarding the topic of the assertion (or riddle). Through these explanations, the robot either informed children about its lack of human psychological capacities and machine status (transparent condition) or provided children with some neutral information about itself that did not address these issues (control condition). For instance, after children guessed whether the assertion "My favorite color is red"

Fig. 2. Nao robot.

was true or false, the robot would, in the transparent condition, say: “I don’t have a favorite color, or other favorite things. I’m sort of a machine, and machines don’t like or dislike anything.” In the control condition, by contrast, it would say: “I’m red and white myself, but red isn’t my favorite color. I like all colors equally, so I don’t really have a favorite color.”

The content of the explanations in the transparent condition was largely based on the manipulation by Van Straten et al. (2020c) to ensure the programmatic character of the present research and increase its cumulative insights. In the control condition, we aimed for explanations that were as similar as possible to the explanations in the transparent condition, but without addressing the robot’s working and technological nature. For instance, in the transparent condition, the robot explained that its machine nature is the reason for its lack of preference (see example provided above). In the control condition, the robot expressed the same lack of preference, but without explaining its cause. This differs from the approach by Van Straten et al. (2020), who, in their control condition, had the experimenter offer information that was completely unrelated to the robot. Making the information across conditions as similar as possible increased the precision of our manipulation. It could be argued that the information in the control condition appears to give the robot a sense of personality. However, the same applies to the majority of CRI scenarios in which a robot is untransparent about its limitations. Contrasting the transparent condition with one that closely mirrors how robots are usually portrayed to children thus increases the relevance of our findings.

The explanations were matched in length across the experimental conditions (i.e., the length of the individual explanations did not differ from each other by more than 5 characters excl. spaces). Appendix A contains all English translations of the information provided by the robot in each of the conditions.

3.3. Procedure

Three female experimenters alternated in conducting the experiment, which took place in a room separated from the main hall of the museum by glass doors. The robot was activated before children entered the experimental room. Upon their entrance, the experimenter introduced herself to the participant and their company. After parents had read an information letter about the study, they were asked to give consent via a secured online consent form.. Children were given an information sheet that explained, in child-appropriate language, what participation entailed, that their data would be analyzed in pseudonymized form, and that participation could be stopped at any time without providing a reason. If necessary, the experimenter or a parent read the information to the children.

The museum asked us to allow children’s parents and other company to be present during the experiment. Thus, the experimenter explained to them that they could stay in the experimental room if they wished, provided they would not interfere with the experiment (we controlled for this in our analyses, see Section 3.5). Once they had left or taken a seat, the experimenter asked whether the child understood the information sheet. Remaining questions were answered unless this interfered with the experimental purposes. If so, the question was postponed until after the debriefing. Then, the child was asked to sit down in front of the robot at a distance they felt comfortable with. At this point, the experimenter reiterated that participation could always be stopped and asked the child whether they would still like to participate. If so, the experiment began.

The study relied upon the WOZ approach, in which an experimenter manually controls the robot without participants being aware of this (see Riek, 2012, for a review of HRI studies that also used this set-up). After the child had indicated that they wanted to begin the interaction, the experimenter took place behind a laptop to control the robot. Both the experimenter and the child’s company were seated behind the child to minimize distractions and the salience of the WOZ procedure (see Fig. 3 for a depiction of the experimental set-up). When the child directed questions or comments toward the experimenter during their interaction with the robot, the experimenter responded as briefly and neutrally as possible. When the interaction was finished, she asked the child to join her at a table to administer the questionnaire. Here, too, children were seated with their backs toward their company (see Fig. 3). Following the approach by Leite and Lehman (2016), the experimenter familiarized children with the question format and answer scale using several practice items (e.g., “I like candy”, “I like Brussel’s sprouts”). Once the child had indicated to understand the procedure, the questionnaire was administered. The entire questionnaire procedure took 6 min on average.

After finishing the questionnaire, the child watched a debriefing video on a tablet. In the video, we explained to children how robots differ from humans (e.g., that they cannot think for themselves or experience emotions); revealed the WOZ approach and the pre-programmed nature of the interaction; and explained the experimental manipulation and its purpose in child-appropriate language. Some of the

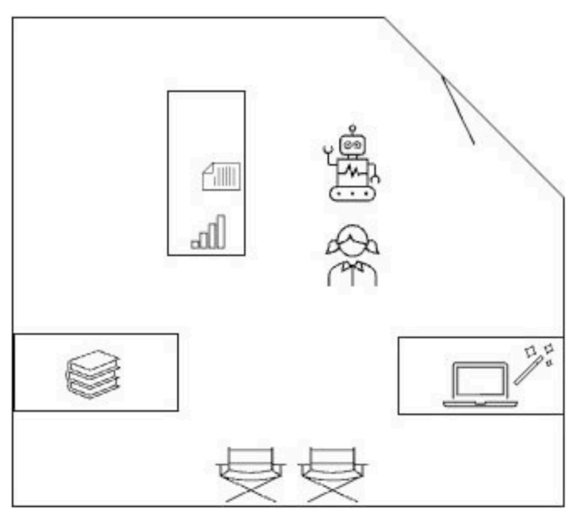

Fig. 3. Visualization of the experimental set-up.
Note. The experimenter (i.e., “wizard”) was seated behind the laptop. Children’s company was seated in the chairs or at the reading table. During the questionnaire, children sat with their back toward their company (i.e., facing the answering scale). The experimenter sat at the other side of the table.

information was not new to children who had been exposed to the transparent condition. However, we wanted to ensure that all children had received this information before leaving the experimental room. When the video was finished, children's and/or parents' remaining questions were answered. Children were asked whether they wanted to take a picture with the robot before they left.

3.4. Measures

The first page of the questionnaire contained demographic questions that children were asked to fill in themselves: Two questions asked children about their age (i.e., 8, 9, or 10) and gender (i.e., boy/girl), and an open question asked them to write down the month of their birthday (i.e., to obtain a more precise age measurement). The rest of the questions were orally administered by the experimenter, and all used the same visualized 5-point Likert scale (adapted from Severson and Lemm, 2016). The scale's verbal answer options ran from "does not apply at all" to "applies completely" and were accompanied by bars of increasing height that clarified their meaning without providing an indication as to the valence of the answer options (e.g., through colors or smileys). The same answer scale has successfully been used in earlier data collections with children in a similar age range (e.g., Van Straten et al., 2020c).

The questionnaire started with a measure of anthropomorphism, followed by measures of perceived similarity, closeness, and trust, and ended with a treatment check. The measures were ordered such that the ones administered earlier would minimally influence the ones thereafter. The one-factorial structure of the measures (except the treatment check) was confirmed in earlier data collections (e.g., Van Straten et al., 2020c). The indicators per concept as well as the response scale used in the present study can be consulted in Van Straten et al. (2020c).

3.4.1. Treatment check

The treatment check consisted of six items (i.e., three per experimental condition) that asked children whether the robot had told them certain things during the interaction. Items one, three, and four referred to information that only children who had been exposed to the transparent condition had received. Items two, five, and six contained information that was only present in the other experimental condition. The items were ordered such that children would not have to indicate three times in a row that they did not recognize the information that an item contained, which may have caused discomfort. In addition, the ordering of the items made it difficult to identify the expected answer pattern.

The items referring to information provided in the transparent condition loaded onto one factor that explained 72% of the variance ($\alpha = 0.88$). An index score was computed by averaging the items ($M = 3.13$, $SD = 1.57$, skewness $= -0.082$, kurtosis $= -1.638$). Similarly, the items referring to the condition in which the robot did not provide transparent information loaded onto one factor that explained 71% of the variance ($\alpha = 0.88$). An index score was created by averaging the items ($M = 3.21$, $SD = 1.50$, skewness $= -0.194$, kurtosis $= -1.482$).

3.4.2. Anthropomorphism

We assessed anthropomorphism using a four-item scale based on the technology dimension of the Individual Differences in Anthropomorphism Questionnaire-Child Form (IDAQ-CF) by Severson and Lemm (2016). In an earlier study (Van Straten, 2020c), we used three measures of anthropomorphism (i.e., IDAQ-CF, Godspeed, and visualized semantic differentials) to assess whether findings on one particular measure would not simply result from a close match between its indicators and the experimental manipulation. As our findings were consistent across all measures (see Van Straten et al., 2020c) and given the similarity between the current manipulation and the one used in Van Straten et al. (2020c), we only used one measure (i.e., the multi-item scale with the best psychometric properties) in the current study.

The one-factorial structure of the scale (principal axis factoring, direct oblimin rotation; this type of analysis was used for all scales) explained 39% of the variance. The scale's internal consistency was acceptable ($\alpha = 0.70$) and could not be substantially improved by removing an item. We averaged the items to create an index score of anthropomorphism ($M = 2.65$, $SD = 0.95$, skewness $= 0.253$, kurtosis $= -0.645$).

3.4.3. Perceived similarity

To assess children's perception of the robot's similarity to themselves, we administered a four-item scale based on the attitude dimension of McCroskey et al. (1975) perceived homophily measure. The scale had a one-factorial structure that explained 41% of the variance, and internal consistency was acceptable ($\alpha = 0.73$). An index score of perceived similarity was computed by averaging the items ($M = 2.02$, $SD = 0.72$, skewness $= 0.262$, kurtosis $= -0.732$).

3.4.4. Closeness

Closeness was measured through a five-item scale that we developed for CRI research and that we have validated among children aged 7 to 11 (Van Straten et al., 2020a). The one-factorial structure of the scale explained 51% of the variance ($\alpha = 0.84$). Averaging the items resulted in an index score of closeness ($M = 3.42$, $SD = 0.72$, skewness $= -0.290$, kurtosis $= 0.422$).

3.4.5. Trust

We assessed trust through a four-item scale based on Larzelere and Huston's (1980) measure of the concept. The scale had a one-factorial structure that explained 52% of the variance ($\alpha = 0.79$). We averaged the items to create an index score of trust ($M = 4.06$, $SD = 0.72$, skewness $= -0.737$, kurtosis $= 0.150$).

3.5. Analytical approach

We analyzed the data using SPSS Statistics (Version 28). The data were considered to be normally distributed when skewness and kurtosis ranged between −2 and 2 (George and Mallery, 2010). This was the case for all dependent variables, including the treatment check. The treatment check as well as H1 and H2 were tested through a series of separate ANOVAs with the robot's provision of information about its lack of human psychological capacities and machine status as the independent variable. Hypotheses 3 and 4 were tested using Hayes' PROCESS macro (Version 4.2), with both mediators entered in the same analysis (model 4, 5000 bootstrapped samples).

The assumption of homoscedasticity was met for the measures of perceived similarity and closeness, but not for anthropomorphism, trust, and the treatment check. Therefore, we additionally performed the Welch test and consulted the parameter estimates with robust standard error (using the heteroscedasticity-consistent standard error HC3; Hausman and Palmer, 2012) for the relevant ANOVAs. We also ran the PROCESS models once with and once without heteroskedasticity-consistent inference (HC3). As the pattern of results was consistent across all the analyses, we report the outcomes of the ANOVAs and regular PROCESS models. We initially controlled, in all analyses and with dummy variables, for a) the experimenter conducting the research; b) the presence of adults (other than the experimenter); c) their interference with the experimental procedure; d) mistakes made in controlling the robot (split into two dummy variables representing either mistakes due to which children missed one robot answer or other mistakes such as timing); and e) the occurrence of robot malfunctions. As the pattern of results of the models with and without control variables was consistent, we report the latter.

4. Results

4.1. Treatment check

Children exposed to the transparent condition indicated more often that the robot had provided them with transparent information (M = 4.60, SD = 0.52) than did children exposed to the other experimental condition (M = 1.74, SD = 0.74). This difference was significant, F(1, 273) = 1359.644, $p < .001$, $\eta^2 = 0.833$. In contrast, children to whom the robot did not talk about its lack of human psychological capacities and machine status more often indicated that they recognized things that the robot only mentioned to them (M = 4.48, SD = 0.65) than children in the transparent condition (M = 1.87, SD = 0.83). This difference was also significant, F(1, 273) = 856.344, $p < .001$, $\eta^2 = 0.758$. Thus, the treatment check was successful.

4.2. Tests of hypotheses

As predicted in H1a and H1b, children in the transparent condition experienced less closeness toward and trust in the robot than did children who had not received transparent information (see Table 1 for all statistics relating to H1 and H2). Therefore, H1 is supported. As specified in H2a and H2b, children in the transparent condition anthropomorphized the robot less and perceived the robot as less similar to themselves than did children who had not received transparent information. Our findings thus support H2.

Hypothesis 3 posited that a decrease in children's tendency to anthropomorphize the robot and their perception of the robot's similarity to themselves would be associated with decreased feelings of closeness toward (H3a) and trust in (H3b) the robot. In terms of H3a, the relationships between anthropomorphism and closeness, b = 0.164, SE = 0.059, p = .006, [95% CI 0.048; 0.281], and perceived similarity and closeness, b = 0.138, SE = 0.061, p = .025, [95% CI 0.018; 0.259], were significant. As to H3b, the relationship between anthropomorphism and trust was significant, b = 0.198, SE = 0.060, p = .001, [95% CI 0.080; 0.316]. However, the relationship between perceived similarity and trust was not significant, b = 0.060, SE = 0.062, p = .332, [95% CI −0.062; 0.182]. In sum, H3a is supported while H3b is partially supported.

Finally, H4 stated that the negative effects of the robot's provision of information about its lack of human psychological capacities and machine status on children's feelings of closeness toward and trust in the robot would be transmitted by children's anthropomorphic tendencies and their perception of the robot's similarity to themselves. In line with this hypothesis, the negative effects of the manipulation on closeness through anthropomorphism, b = −0.205, bootstrapped (bt) SE = 0.089 [95% bt CI −0.378; −0.033], and on closeness through perceived similarity, b = −0.058, bt SE = 0.031 [95% bt CI −0.124; −0.002], were significant. The manipulation also had a significant negative effect on trust through anthropomorphism, b = −0.248, bt SE = 0.081 [95% bt CI −0.407; −0.088]. However, the indirect effect on trust through perceived similarity was not significant, b = −0.025, bt SE = 0.029 [95% bt CI −0.087; 0.029]. The direct effects of our manipulation on closeness and trust that were found in the ANOVAs were no longer significant in the mediation model (b = 0.057, SE = 0.113, p = .615 [95% CI −0.165; 0.278] for closeness; b = 0.081, SE = 0.114, p = .480 [95% CI −0.144; 0.305] for trust). In sum, the results partially support H4.

5. Discussion

The present study investigated whether a robot's provision of information about its lack of human psychological capacities and machine status influences children's perceptions of, and relationship formation with, a social robot. When the robot provided such transparent information about itself, children's tendency to anthropomorphize the robot and their perception of the robot's similarity to themselves decreased, as did children's feelings of closeness toward and trust in the robot. Anthropomorphism and perceived similarity were positively associated with closeness, and anthropomorphism was also positively related to trust. Accordingly, our results show that both anthropomorphism and perceived similarity mediated the effect of transparent information on closeness, while only anthropomorphism mediated its effect on trust. As the psychological mechanisms of child-robot relationship formation remain underexplored (Van Straten et al., 2020b) our findings shed some first light on the role that children's perceptions of a social robot play in the emergence of child-robot relationships. Our study also shows that a social robot can influence how children perceive and relate to it by telling them about its lack of human psychological capacities.

Our results add to CRI research on the effects of transparency about social robots in two ways. First, our findings reconfirm that, next to transparency about robots' teleoperated working (Cameron et al., 2017; Chernyak and Gary, 2016; Somanader et al., 2011; Tozadore et al., 2017; Van Straten et al., 2022), transparent information about robots' lack of human psychological capacities can alter children's responses to robots (see also Van Straten et al., 2020c). In the future, robots may increasingly function autonomously (e.g., Stapels and Eyssel, 2021) and be equipped with more sophistcated capacities. As a result, our understanding of 'humanness' and the basis on which we distinguish between humans and machines may change, possibly forcing us to reconceptualize what 'transparent information' is (see Festerling and Siraj, 2022). Yet, even if the autonomy and capacities of future robots increase, differences between humans and robots will remain (e.g., Fox and Gambino, 2021). Transparency about robots' lack of human capacities will thus continue to be relevant.

Second, our study shows that robots themselves can effectively convey transparent information to children. Transparency thus does not require adult intervention. This is an important finding because children's encounters with future, more autonomous robots are probably not always supervised by adults. Overall, our results do not only support the relevance of normative calls for transparency about social robots, but also extend prior empirical research on the topic (e.g., Cameron et al., 2017; Tozadore et al., 2017; Van Straten et al., 2020c;2022) to likely scenarios of CRI in the future. In line with Schött et al.'s (2023) aforementioned distinction between transparency *on* and *through* robots, our findings also suggest that future studies in CRI should look more into the effects of transparency through robots, in which external explanations become unnecessary because robots are inherently transparent through their communication, actions, and design.

Our study also relates to research on explainable robotics. Explainable robotics aims to increase robots' predictability and understandability and decrease people's overapplication of mental models from interpersonal communication to HRI, by having robots explain their

Table 1
Tests of between-subjects effects of one-way ANOVA.

				Transparent		Control	
Dependent variable (df)	F	p	η^2	M	SD	M	SD
Closeness (1, 274)	5.891	.016	.02	3.32	0.72	3.53	0.70
Trust (1, 274)	5.581	.019	.02	3.96	0.80	4.16	0.64
Anthropomorphism (1, 273)	207.508	<0.001	.43	2.02	0.63	3.27	0.79
Perceived similarity (1, 273)	25.298	<0.001	.09	1.80	0.70	2.23	0.68

actions (see, e.g., De Graaf et al., 2021). Prior studies have shown that robots that explain what they (do not) do increase users' understanding and evaluation of them (e.g., Stange and Kopp, 2020; see more generally also Schött et al., 2023). Qualitative evidence suggests that people find it especially important that robots explain *why* they act as they do (Han et al., 2021). While, generally, such explanations can change people's thoughts about robots (De Graaf et al., 2021), *transparent* explanations may be particularly effective to this end for user groups with strong anthropomorphic tendences, such as children (see, e.g., Epley et al., 2007).

The decrease in trust found in the present study clarifies the findings by Van Straten et al. (2020c), who discussed the possibility that children trusted the robot less because an adult, rather than the robot itself, provided the transparent information. Consequently, children may have gotten the impression that the robot withheld this information from them (van Straten et al., 2020c). However, our finding that children's trust in the robot also decreased when the robot provided this information itself confirms that it is the *nature* of the information rather than its *source* that caused the decrease in children's trust. Next to the decrease in trust, transparent information reduced children's closeness in the present study but not in Van Straten et al. (2020c). This new finding, however, does not unequivocally show that children considered the transparent information to be *more* credible when provided by the robot. The two studies differed in multiple ways (i.e., in terms of age range, setting, interaction task). Therefore, multiple explanations for the different findings are conceivable.

Our results support Scheutz's (2012) argument for being open about a robot's machine status and (lack of) capacities and calls for responsible robotics more broadly (i.e., Boden et al., 2017; Schaake, 2021; Université de Montréal, 2018). Indeed, transparency seems to be an effective means to purposefully influence how children perceive and relate to a social robot. Scheutz (2012, p. 218) has argued that a social robot should, through its appearance and behavior, continuously signal "that it is a machine, that it does not have emotions, [and] that it cannot reciprocate […]", as this may decrease the likelihood that people may form emotional attachments to the robot. In our study, this continuous signaling was achieved by having the robot explain to children what it was (not) capable of and how it works throughout the interaction. Thus, in the transparent condition more so than in the control condition, the robot disrupted the conversational flow by immediately demystifying the interaction (and itself) to children. Rather than a confounding factor, we consider this an inevitable consequence of establishing continuous transparency as requested by, among others, Scheutz (2012).

The very strong effect of transparency on children's anthropomorphic tendencies may in part result from the similarity between the transparent information that the robot shared about itself and the content of the items that we used to measure anthropomorphism. Still, this measure cannot simply be considered a manipulation check, as it asks children about their *perception* of the robot – which may be influenced by, but does not per definition equal, the information they received during the experiment. Moreover, the information provided by the robot was meant to make children aware of its lack of human psychological capacities. Conceptual overlap between this information and items used to measure anthropomorphism is therefore to some extent inevitable.

In contrast to the effect of transparent information on anthropomorphism, the effects on closeness, trust, and perceived similarity were significant but small. While transparency can alter children's perceptions of, and relationship formation with, a social robot, it should thus not be expected to prevent child-robot relationship formation altogether. More generally, transparent information may primarily be effective in changing children's reasoning about those robot characteristics that are explicitly addressed. To adults it may seem obvious that a robot that cannot think like a human and does not have emotions can barely be a friend, but (young) children may not be expected to make such inferences, given their strong imagination. As a consequence, if the aim is to purposefully decrease, for example, children's expectations of robots' friendship potential, it may be advisable to directly address the boundaries of human-robot relationships in the transparent information being provided.

Despite concerns about the emergence of social relationships between humans and robots, some scholars have argued that such relationships may be part of a more positive overall process. Collins (2017), for example, has criticized the 'empty' use of terms like deception and argues that social robots should be used in the most efficient way possible. Similarly, Pearson and Borenstein (2014) have argued that child-robot relationship formation is, to some degree, required for robots to effectively contribute to children's welfare. These views raise the question of whether transparency about social robots is desirable at all, given its negative effect on children's tendency to anthropomorphize the robot and their perception of its similarity to themselves and, by extension, their feelings of closeness toward and trust in the robot.

Against this background, Sandry (2015, p. 10), for example, has maintained that anthropomorphic responses to social robots are particularly valuable when "roboticists do not try to reinforce [these responses, and when] they are tempered by a parallel clarity of understanding the robot as a machine." Likewise, Diaz et al. (2011) have posited that encouraging children to form realistic expectations of social robots may facilitate the emergence and maintenance of long-term child-robot relationships (see also Caudwell et al., 2019, on the development of durable human-robot relationships). Moreover, people do not establish relationships in single encounters as the development of closeness and trust normally takes time (e.g., Berscheid and Regan, 2005). Thus, children's decreased feelings of closeness toward, and trust in, the robot as found in the present study may be interpreted as a form of natural prudence, which may – in the long run – benefit rather than harm the societal potential of CRI applications.

In contrast to our expectations, perceived similarity did not mediate the effect of transparent information on children's trust in the robot. This finding may indicate that children's trust in a robot is independent of their perception of the robot's similarity to themselves. Possibly, children's awareness of a robot's lack of human psychological capacities primarily reduces children's trust in a robot's *competencies*, which may partly underlie but does not coincide with, a more 'social' kind of trust in the robot's benevolence and honesty (for a meta-analysis on social vs. competency trust in CRI, see Stower et al., 2021; for qualitative evidence for the interrelatedness of both kinds of trust in CRI, see Van Straten et al., 2018). Accordingly, children's perception of the robot's similarity to themselves may not have mediated their level of trust in the robot because perceived similarity predicts *social* attraction (see, e.g. Montoya et al., 2008), and, by extension, *social* rather than *competency* trust.

Our study has at least five limitations. First, our study was conducted in a museum. As a consequence, the experimental procedure could be controlled somewhat less than in a lab, notably when parents were present. Second, our mediating variables anthropomorphism and perceived similarity were measured rather than manipulated. As outlined in the theory section, there are theoretical reasons to expect anthropomorphism and perceived similarity to predict children's closeness and trust toward the robot. Still, we cannot empirically preclude the opposite causal direction, nor can we rule out threats to the internal validity of the association between the two mediators and the dependent variables. Third, some of the analyses we performed were underpowered (i.e., the analyses of the main effect on closeness and the mediating effects of anthropomorphism) and a replication with adequately powered samples seems desirable.

Fourth, we only used self-report measures to assess children's robot perceptions and child-robot relationship formation as this is more straightforward and less time-consuming than the use of observational measures. Although our measures were validated and successfully used in prior CRI studies (e.g., Van Straten et al., 2020c, 2022), they may be subject to cognitive and social-desirability biases. Fifth and finally, our study relies upon a sample of children within a specific age range, which limits the generalizability of our findings. While younger children may for instance more readily trust and anthropomorphize robots (e.g., Di Dio et al., 2020; Manzi et al., 2020), older children may have more critical stances toward social robots to begin with (e.g., Kahn et al., 2012).

Apart from opportunities for future studies that follow from the aforementioned limitations of our own work, we have three directions for future research on the effects of being transparent to children about social robots. First, longitudinal research should investigate how transparency affects children's perception of, and relationship formation with, social robots in the long term. In our studies, children could only interact with the robot once and for a short amount of time and our findings illustrate the influences of a transparent approach on children's *initial* responses to social robots only. Once the novelty effect wears off (see, e.g. Leite et al., 2013), children may become more critical of a robot and, by consequence, more open to transparent information about its limitations. Second, it would be useful to explore for how long transparent information offered during a particular encounter retains its influence on children's thoughts and feelings. We found relatively small effects of transparency on closeness and trust, which may indicate that transparent information needs to be provided repeatedly for its influence to become stronger and to be persistent over time. Third, future studies should aim to more generally explore how a robot's contributions to an interaction may influence children's thoughts and feelings. Investigating such influences could not only expand our knowledge about the proactive role that robots may play in the development of child-robot relationships but would also be relevant to CRI research more generally.

Author contributions

Authors: Caroline L. van Straten, Jochen Peter, & Rinaldo Kühne

Caroline L. van Straten: conceptualization, methodology, data collection, analysis, writing (original draft preparation), writing (review & editing)

Jochen Peter: conceptualization, methodology, analysis, writing (review & editing), funding acquisition

Rinaldo Kühne: conceptualization, methodology, analysis, writing (review & editing)

Declaration of Competing Interest

The authors declare to have no competing interests.

Data availability

The data will be made available on OSF (osf.io/3867k).

Acknowledgements

We would like to thank NEMO Science Museum Amsterdam for the opportunity to conduct our study at their R&D Lab, as well as for their assistance during data collection. In addition, we thank all (parents of) children who participated in this study.

Funding statement

This work was supported by the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation program (grant agreement No. [682733]) to the second author.

Appendix A: English translations of the information provided by the robot

	Transparent information	Non-transparent information
1	There are many robots exactly like me, and those are all called Nao, too! If you replaced me with a different Nao-robot, you wouldn't notice that. Just like when you switch between computers: You don't notice that either, because every computer does exactly the same.	My name is indeed Nao. I have talked with many children like you, but I have never met a child who had the same name as I do. Sometimes children ask me what my name means, but I actually don't precisely know that. Perhaps, my name doesn't mean anything at all.
2	I have a computer in my head, with many computer programs. In those programs, you can change the language I speak. You can, for instance, program me to talk to you in English or French. But I can't choose myself which language I speak.	I can, for instance, talk to you in English or French. That's very useful, because it enables me to talk to children from many different countries. But usually I talk with Dutch children, so I don't speak the other languages very often.
3	I don't have a favorite color, or other favorite things. I'm sort of a machine, and machines don't like or dislike anything.	I'm red and white myself, but red isn't my favorite color. I like all colors equally, so I don't really have a favorite color.
4	I can tell this riddle because it's in my computer program. I can't come up with riddles on my own because I can't think on my own. I'm just like a computer: they can't come up with funny things, either!	I can tell some other jokes but those are all quite similar to this one, because the solution to each of them is a silly fantasy word. I learned the jokes I know from the researchers.
5	If I lose a game, I don't care. Because I'm a machine, and machines don't notice it when they lose. Just like a computer doesn't realize when it loses a game against you! In fact, I'm never happy, angry, or sad, because I consist of plastic and wires. So, I don't feel anything at all!	I don't mind losing a game. With most games you mostly just need to be lucky anyway, such that you can't help it if you lose. So, that's no big deal then! You can't lose at all at this game, by the way. I know some children don't like to lose, so that's convenient.
6	Children often hope that other children will like them. But I'm a robot and can't think for myself. Therefore, I also can't think about what people may think of me or realize that something I said earlier was in fact not very smart to say. I just do everything automatically.	I actually never think about what people may think of me. I simply am who I am. Most children like to play with me, but if they don't like to do so, I don't mind, either. After all, I can't help it whether children like to do something or not.
7	Some children want to become teachers when they grow up. In a few years, you will be a lot more grown up than you are right now, and you may then want to do different things than you do now. I will never change and will always stay exactly the same. That's no problem because just like machines, robots do what people want them to do, but don't want anything themselves.	Some children want to become teachers when they grow up. I know that teachers are very smart, because when I visited a primary school, I saw that they teach children many different things. Maybe I can teach children some things too, like math or grammar, but I'm not an actual teacher, of course.
8	I say exactly the same things to all the children I play with. That's because I can't really react to what people do or say. I just wait until it's my turn to speak, and then I say precisely what is in my computer.	I say exactly the same things to all the children I play with because I play the same game with each of them. I can also pose different questions to everyone, but many different questions would be needed for that.

References

Atherton, G., Cross, L., 2018. Seeing more than human: autism and anthropomorphic theory of mind. Front. Psychol. 9, 1–18. https://doi.org/10.3389/fpsyg.2018.00528.

Bauminger-Zviely, N., Agam-Ben-Artzi, G., 2014. Young friendship in HFASD and typical development: friend versus non-friend comparisons. J. Autism Dev. Disord. 44 (7), 1733–1748. https://doi.org/10.1007/s10803-014-2052-7.

Berger, C.R., Calabrese, R.J., 1975. Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. Hum. Commun. Res. 1 (2), 99–112. https://doi.org/10.1111/j.1468-2958.1975.tb00258.x.

Berscheid, E., Regan, P., 2005. The Psychology of Interpersonal Relationships. Pearson Education, Upper Saddle River, NJ.

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B., Winfield, A., 2017. Principles of robotics: regulating robots in the real world. Connect. Sci. 29 (2), 124–129. https://doi.org/10.1080/09540091.2016.1271400.

Borenstein, J., Pearson, Y., 2013. Companion robots and the emotional development of children. Law, Innov. Technol. 5 (2), 172–189. https://doi.org/10.5235/17579961.5.2.172.

Bumby, K., & Dautenhahn, K., 1999. Investigating children's attitudes towards robots: a case study. Proceedings of the Third International Cognitive Technology Conference. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.2906&rep=rep1&type=pdf.

Cameron, D., Fernando, S., Collins, E.C., Millings, A., Szollosy, M., Moore, R., Sharkey, A. J.C., & Prescott, T., 2017. You made him be alive: children's perceptions of animacy in a humanoid robot. Proceedings of the Conference on Biomimetic and Biohybrid Systems, 73–85. https://doi.org/10.1007/978-3-319-63537-8_7.

Caudwell, C., Lacey, C., & Sandoval, E.B., 2019. The (ir)relevance of robot cuteness: an exploratory study of emotionally durable robot design. Proceedings of the 31st Australian Conference on Human-Computer Interaction, 64–72. https://doi.org/10.1145/3369457.3369463.

Chernyak, N., Gary, H.E., 2016. Children's cognitive and behavioral reactions to an autonomous versus controlled social robot dog. Early Educ. Dev. 27 (8), 1175–1189. https://doi.org/10.1080/10409289.2016.1158611.

Cole, M., Cole, S., Lightfoot, C., 2005. The Development of Children, 5th Ed. Worth.

Collins, E.C., 2017. Vulnerable users: deceptive robotics. Connect. Sci. 29 (3), 223–229. https://doi.org/10.1080/09540091.2016.1274959.

De Graaf, M.M., Dragan, A., Malle, B.F., Ziemke, T., 2021. Introduction to the special issue on explainable robotic systems. ACM Transactions on Human-Robot Interaction 10 (3), 1–4. https://doi.org/10.1145/3461597.

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. Front. Psychol. 11 https://doi.org/10.3389/fpsyg.2020.00469.

Díaz, M., Nuño, N., Saez-Pons, J., Pardo, D.E., & Angulo, C. (2011). Building up child-robot relationship for therapeutic purposes: from initial attraction towards long-term social engagement. Proceedings of the International Conference on Automatic Face and Gesture Recognition, 927–932. https://doi.org/10.1109/FG.2011.5771375.

Edwards, A., Edwards, C., Spence, P.R., Harris, C., Gambino, A., 2016. Robots in the classroom: differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher. Comput. Hum. Behav. 65, 627–634. https://doi.org/10.1016/j.chb.2016.06.005.

Emmeche, C., 2014. Robot friendship: can a robot be a friend? Int. J. Signs Semiotic Syst. 3 (2), 26–42. https://doi.org/10.4018/IJSSS.2014070103.

Epley, N., Waytz, A., Cacioppo, J.T., 2007. On seeing human: a three-factor theory of anthropomorphism. Psychol. Rev. 114 (4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864.

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. Behav. Res. Methods 39, 175–191. https://doi.org/10.3758/bf03193146.

Festerling, J., Siraj, I., 2022. Anthropomorphizing technology: a conceptual review of anthropomorphism research and how it relates to children's engagements with digital voice assistants. Integr. Psychol. Behav. 56 (3), 709–738. https://doi.org/10.1007/s12124-021-09668-y.

Fox, J., Gambino, A., 2021. Relationship development with humanoid social robots: applying interpersonal theories to human/robot interaction. Cyberpsych. Beh. Soc. N. https://doi.org/10.1089/cyber.2020.0181.

George, D., Mallery, P., 2010. SPSS For Windows step By step: A simple Guide and reference. 17.0 Update, 10th ed. Pearson, Boston, MA.

Gleason, T.R., 2013. Imaginary Relationships. The Oxford Handbook of the Development of Imagination.

Han, Z., Phillips, E., Yanco, H.A., 2021. The need for verbal robot explanations and how people would like a robot to explain itself. ACM Trans. Human-Robot Interact. 10 (4), 1–42. https://doi.org/10.1145/3469652.

Hancock, P.A., Kessler, T.T., Kaplan, A.D., Brill, J.C., Szalma, J.L., 2020. Evolving trust in robots: specification through sequential and comparative meta-analyses. Hum. Factors 63 (7), 1196–1229. https://doi.org/10.1177/0018720820922080.

Hartup, W.W., Vangelisti, A.L., Perlman, D., 2006. Relationships in early and middle childhood. The Cambridge Handbook of Personal Relationships. Cambridge University Press.

Hausman, J., Palmer, C., 2012. Heteroskedasticity-robust inference in finite samples. Econ. Lett. 116 (2), 232–235. https://doi.org/10.1016/j.econlet.2012.02.007.

Hubbard, F.P., 2011. Do androids dream?": personhood and intelligent artifacts. Temple Law Rev 83 (2), 405–473. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/temple83&section=16.

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

Keller, J., 1997. Autonomy, relationality, and feminist ethics. Hypatia 12 (2), 152–164. https://doi.org/10.1111/j.1527-2001.1997.tb00024.x.

Kim, K.J., Park, E., Shyam Sundar, S., 2013. Caregiving role in human-robot interaction: a study of the mediating effects of perceived benefit and social presence. Comput. Hum. Behav. 29 (4), 1799–1806. https://doi.org/10.1016/j.chb.2013.02.009.

Kory Westlund, J. M., Martinez, M., Archie, M., Das, M., & Breazeal, C.L., 2016. Effects of framing a robot as a social agent or as a machine on children's social behavior. Proceedings of the 25th International Symposium on Robot and Human Interactive Communication, 688–693. https://doi.org/10.1109/ROMAN.2016.7745193.

Larzelere, R.E., Huston, T.L., 1980. The dyadic trust scale: toward understanding interpersonal trust in close relationships. J Marriage Fam 42 (3), 595–604. https://doi.org/10.2307/351903.

Lee, K.M., Jung, Y., Kim, J., Kim, S.R., 2006a. Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. Int. J. Hum.-Comput. St. 64 (10), 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002.

Lee, K.M., Peng, W., Jin, S.A., Yan, C., 2006b. Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. J. Commun. 56 (4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x.

Leite, I., & Lehman, J.F., 2016. The robot who knew too much: toward understanding the privacy/personalization trade-off in child-robot conversation. Proceedings of the 15th Conference on Interaction Design and Children, 379–387. https://doi.org/10.1145/2930674.2930687.

Leite, I., Martinho, C., Paiva, A., 2013. Social robots for long-term interaction: a survey. Int. J. Soc. Robot. 5 (2), 291–308. https://doi.org/10.1007/s12369-013-0178-y.

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., Marchetti, A., 2020. A robot is not worth another: exploring children's mental state attribution to different humanoid robots. Front. Psychol. 11 https://doi.org/10.3389/fpsyg.2020.02011.

McCroskey, J.C., Richmond, V.P., Daly, J.A., 1975. The development of a measure of perceived homophily in interpersonal communication. Hum. Commun. Res. 1 (4), 323–332. https://doi.org/10.1111/j.1468-2958.1975.tb00281.x.

Montoya, R.M., Horton, R.S., Kirchner, J., 2008. Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. J. Soc. Pers. Relat. 25 (6) https://doi.org/10.1177/0265407508096700.

O'Keefe, D.J., 2002. Persuasion: Theory and Research, 2nd ed. Sage Publications.

Oranç, C., Küntay, A.C., 2020. Children's perception of social robots as a source of information across different domains of knowledge. Cognitive Dev 54 (100875). https://doi.org/10.1016/j.cogdev.2020.100875.

Paiva, A., Mascarenhas, S., Petisca, S., Correia, F., Alves-Oliveira, P., 2018. Towards more humane machines: creating emotional social robots. In: Da Silva, S.G. (Ed.), New Interdisciplinary Landscapes in Morality and Emotion. Routledge, pp. 125–139.

Pashevich, E., 2021. Can communication with social robots influence how children develop empathy? Best-evidence synthesis. AI Soc. 37, 579–589. https://doi.org/10.1007/s00146-021-01214-z.

Pearson, Y., Borenstein, J., 2014. Creating "companions" for children: the ethics of designing esthetic features for robots. AI Soc 29 (1), 23–31. https://doi.org/10.1007/s00146-012-0431-1.

Petrina, N., Carter, M., Stephenson, J., 2014. The nature of friendship in children with autism spectrum disorders: a systematic review. Res. Autism Spect. Dis. 8 (2), 111–126. https://doi.org/10.1016/j.rasd.2013.10.016.

Prescott, T.J., Robillard, J.M., 2021. Are friends electric? The benefits and risks of human-robot relationships. IScience 24 (1), 101993. https://doi.org/10.1016/j.isci.2020.101993.

Rieh, S.Y., 2016. Credibility and cognitive authority of information. Encyclopedia of Library and Information Sciences, 3rd ed. CRC Press, pp. 1337–1344. https://doi.org/10.1081/e-elis3-120044103.

Riek, L.D., 2012. Wizard of oz studies in HRI: a systematic review and new reporting guidelines. J Human-Robot Interaction 1 (1), 119–136. https://doi.org/10.5898/jhri.1.1.riek.

Sandry, E., 2015. *Re*-evaluating the form and communication of social robots: the benefits of collaborating with machinelike robots. Int. J. Soc. Robot. 7 (3), 335–346. https://doi.org/10.1007/s12369-014-0278-3.

Schaake, M., 2021. European commission's Artificial Intelligence Act. https://hai.stanford.edu/sites/default/files/2021-06/HAI_Issue-Brief_The-European-Commissions-Artificial-Intelligence-Act.pdf.

Scheutz, M., 2012. The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin, P., Abney, K., Bekey, G.A. (Eds.), Robot ethics: The ethical and Social Implications of Robotics. MIT Press, Cambridge, MA, pp. 205–221.

Schoemann, A.M., Boulton, A.J., Short, S.D., 2017. Determining power and sample size for simple and complex mediation models. Soc. Psychol. Pers. Sci. 8 (4), 379–386. https://doi.org/10.1177/1948550617715068.

Schött, S.Y., Amin, R.M., Butz, A., 2023. A literature survey of how to convey transparency in co-located human–robot interaction. Multimodal Technol. Interact. 7 (3), 25. https://doi.org/10.3390/mti7030025.

Severson, R.L., Carlson, S.M., 2010. Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. Neural Networks 23 (8–9), 1099–1103. https://doi.org/10.1016/j.neunet.2010.08.014.

Severson, R.L., Lemm, K.M., 2016. Kids see human too: adapting an individual differences measure of anthropomorphism for a child sample. J Cogn. Dev. 17 (1), 122–141. https://doi.org/10.1080/15248372.2014.989445.

Sharkey, A.J.C., Sharkey, N., 2020. We need to talk about deception in social robotics! Ethics Inf. Technol. 23, 309–316. https://doi.org/10.1007/s10676-020-09573-9.

Somanader, M.C., Saylor, M.M., Levin, D.T., 2011. Remote control and children's understanding of robots. J. Exp. Child Psychol. 109 (2), 239–247. https://doi.org/10.1016/j.jecp.2011.01.005.

Stange, S., & Kopp, S., 2020. Effects of a social robot's self-explanations on how humans understand and evaluate its behavior. Proceedings of the International Conference on Human-Robot Interaction, 619–627. https://doi.org/10.1145/3319502.3374802.

Stapels, J.G., Eyssel, F., 2021. Robocalypse? Yes, please! The role of robot autonomy in the development of ambivalent attitudes towards robots. Int. J. Soc. Robot. 14 (3), 683–697. https://doi.org/10.1007/s12369-021-00817-2.

Sternberg, R.J., 1987. Liking versus loving: a comparative evaluation of theories. Psychol. B. 102 (3), 331–345. https://doi.org/10.1037/0033-2909.102.3.331.

Stower, R., Calvo-Barajas, N., Castellano, G., Kappas, A., 2021. A meta-analysis on children's trust in social robots. Int. J. Soc. Robot. https://doi.org/10.1007/s12369-020-00736-8.

Tolksdorf, N.F., Siebert, S., Zorn, I., Horwath, I., Rohlfing, K.J., 2020. Ethical considerations of applying robots in kindergarten settings: towards an approach from a macroperspective. Int. J. Soc. Robot. 13 (2), 129–140. https://doi.org/10.1007/s12369-020-00622-3.

Tozadore, D.C., Pinto, A., Romero, R., & Trovato, G., 2017. Wizard of Oz vs autonomous: children's perception changes according to robot's operation condition. Proceedings of the Sixth International Symposium on Robot and Human Interactive Communication, 664–669. https://doi.org/10.1109/ROMAN.2017.8172374.

Tung, F.W., 2016. Child perception of humanoid robot appearance and behavior. Int. J. Hum.-Comput. Int. 32 (6), 493–502. https://doi.org/10.1080/10447318.2016.1172808.

Turkle, S., 2006. A nascent robotics culture: new complicities for companionship. Am. Assoc. Artif. Intell. Tech. Report Series 107–116. https://doi.org/10.4324/9781003074991-12.

Turkle, S., 2007. Authenticity in the age of digital companions. Int. Stud. 8 (3), 501–517. https://doi.org/10.1075/is.8.3.11tur.

Turkle, S., Breazeal, C.L., Dasté, O., Scassellati, B., 2006. First encounters with Kismet and Cog: children respond to relational artifacts. In: Messaris, P., Humphreys, L. (Eds.), Digital media: Transformations in Human Communication. Peter Lang.

Université de Montréal, 2018. *Montreal Declaration*. https://www.montrealdeclaration-responsibleai.com/the-declaration.

Valkenburg, P.M., Piotrowski, J.T., 2017. Plugged in: How media Attract and Affect Youth. Yale University Press.

van Straten, C.L., Peter, J., Kühne, R., de Jong, C., & Barco, A., 2018. Technological and interpersonal trust in child-robot interaction: an exploratory study. Proceedings of the 6th International Conference on Human Agent Interaction, 253–259). https://doi.org/10.1145/3284432.3284440.

van Straten, C.L., Kühne, R., Peter, J., de Jong, C., Barco, A., 2020a. Closeness, trust, and perceived social support in child-robot relationship formation: development and validation of three self-report scales. Int. Stud. 21 (1), 57–84. https://doi.org/10.1075/is.18052.str.

van Straten, C.L., Peter, J., Kühne, R., 2020b. Child–robot relationship formation: a narrative review of empirical research. Int. J. Soc. Robot. 12 (2), 325–344. https://doi.org/10.1007/s12369-019-00569-0.

van Straten, C.L., Peter, J., Kühne, R., Barco, A., 2020c. Transparency about a robot's lack of human psychological capacities: effects on child-robot perception and relationship formation. ACM Trans. Human-Robot Interact. 9 (2) https://doi.org/10.1145/3365668.

van Straten, C.L., Peter, J., Kühne, R., Barco, A., 2022. The wizard and I: how transparent teleoperation and selfdescription (do not) affect children's robot perceptions and childrobot relationship formation. AI Soc. 37 (1), 383–399. https://doi.org/10.1007/s00146-021-01202-3.

van Wynsberghe, A., 2021. Social robots and the risks to reciprocity. AI Soc. 37, 479–485. https://doi.org/10.1007/s00146-021-01207-y.

Vanderborght, M., Jaswal, V.K., 2009. Who knows best? Preschoolers sometimes prefer child informants over adult informants. Infant. Child Dev. 18 (1), 61–71. https://doi.org/10.1002/icd.591.