



When it is ok to give the Robot Less: Children's Fairness Intuitions Towards Robots

Oshrat Ayalon^{1,2} · Hannah Hok³ · Alex Shaw³ · Goren Gordon¹

Accepted: 15 August 2023
© The Author(s) 2023

Abstract

Children develop intuitions about fairness relatively early in development. While we know that children believe other humans care about distributional fairness, considerably less is known about whether they believe other agents, such as robots, do as well. In two experiments ($N = 273$) we investigated 4- to 9-year-old children's intuitions about whether robots would be upset about unfair treatment as human children. Children were told about a scenario in which resources were being split between a human child and a target recipient: either another child or a robot across two conditions. The target recipient (either child or robot) received less than another child. They were then asked to evaluate how fair the distribution was, and whether the target recipient would be upset. Both Experiment 1 and 2 used the same design, but Experiment 2 also included a video demonstrating the robot's mechanistic "robotic" movements. Our results show that children thought it was more fair to share unequally when the disadvantaged recipient was a robot rather than a child (Experiment 1 and 2). Furthermore, children thought that the child would be more upset than the robot (Experiment 2). Finally, we found that this tendency to treat these two conditions differently became stronger with age (Experiment 2). These results suggest that young children treat robots and children similarly in resource allocation tasks, but increasingly differentiate them with age. Specifically, children evaluate inequality as less unfair when the target recipient is a robot, and think that robots will be less angry about inequality.

Keywords Child–robot interaction · Fairness · Social cognitive development · Resource allocation

1 Introduction

Do people think that social robots care about fairness? Until a few years ago, this question might have only been interesting to readers of science fiction. However, as we will review below, social robots are already being used in many situa-

tions in people's daily lives, and in the next decade, they are likely to become a fixture in many classroom contexts [1]. Given that these robots will become more and more integral part of children's lives, it seems important to understand how children think about them. While there is an increasingly large literature on children's thoughts and beliefs about these robots [2–5], and on children's pro-social behavior towards them [6, 7], previous research has overlooked how children think about robots as recipients who may or may not be concerned with distributional inequality and fair treatment in resource distribution. This was the aim of the current experiments, which investigated whether children think that robots would be upset at others having more than them and if their intuitions about these robots would change throughout the course of development. Such research will provide insights into how children think about social robots, as well as their developing intuitions about fairness. Understanding how children think about robots as potential recipients is of theoretical and practical importance. For example, these concerns with fairness are likely to play out in educational settings, where teachers must often assign scarce resources.

✉ Oshrat Ayalon
oayalon@mpi-sws.org
Hannah Hok
hok@uchicago.edu
Alex Shaw
ashaw1@uchicago.edu
Goren Gordon
goren@gorengordon.com

- ¹ Curiosity Lab, Industrial Engineering Department, Tel Aviv University, Tel Aviv, Israel
- ² Max Planck Institute for Software Systems, Saarbrücken, Germany
- ³ Department of Psychology, University of Chicago, Chicago, IL, USA

As people's fairness perceptions about technologies are explored in other contexts, especially in algorithmic decision-making, we highlight the scope of the current study. In many cases, studies exploring fairness in the context of Artificial Intelligence (AI) decision-making are interested in understanding people's perceptions about the extent to which these algorithms are fair or unbiased (e.g., [8, 9]). While the motivations for these studies and the present study share some similarities, the context and scope of the present study differ from AI fairness studies in that we focus on children's perceptions of robots as recipients of fair behavior, not on whether the robots themselves are unfair or biased. Here, given that we are interested in children's fairness perceptions towards robots, the question of whether a technology is perceived as "behaving" in a fair manner is an orthogonal question.

We begin by reviewing previous work on social robots and children's intuitions about fairness. Social robots are designed to use facial expressions, along with verbal and non-verbal communications, to socially interact with people [10, 11]. This aims to facilitate a more natural and fluid integration of robots into our everyday lives, so they may assist in everyday activities. Recent advances in social robotics have made these robots more readily available, bringing us robotic tutors, companions, and peers [12–14]. These robots have been used to assist both older individuals who have cognitive and social impairments [15] and children in learning [16, 17], including commercial products for home use, such as Amazon's "Astro" [18], and robots specified for children's use, such as LuxAI's "QTrobot" for autistic children [19], and "Moxie" from Embodied [20]. Indeed, one of the most promising venues of these social robots is in the field of children's education [21]. Social robots have moved in recent years from the lab into more naturalistic settings, such as schools and even children's homes. They have been used to promote learning in various disciplines, such as language [17, 22], reading [23], and science [24]. Recently, they have been shown to be able to promote children's curiosity [13], creativity [25] and growth mindset [26] via social interactions. These social robots have been used as teachers utilizing several techniques: frontal lecture mode [27], one-on-one interaction [13, 14] and even in small groups [28, 29]. Given children's increased interactions with these social robots, it seems important to understand how children perceive these robots as social entities.

Several studies have examined children's perception of social robots, exploring how they think about the mental states of social robots and moral considerations around how these social robots are treated. Sommer et al. [2] and Di Dio et al. [30] explored children's perceptions about robots compared to other entities, such as human children, both showing that while perceived as less "human" than human-children, they were considered "deserving" of moral considerations. Sommer et al. tested this through 'harming' different enti-

ties by placing a box over them, and measured children's (4- to 10-year-olds) reaction to these actions. These researchers found that children believe that robots deserve some moral treatment, less than living things and more than non-living things [2]. Moreover, the study showed that children's moral concerns about harming certain entities change with age. For non-living entities and animaloid robots, older children thought it was more acceptable to harm them, while their considerations for living entities and humanoid robots remained constant across age [2]. In a different study, Di Dio et al. examined 5- to 6-year-old children's behavior while playing the Ultimatum Game with either a social robot or another child, and measured the perceived mental states of robots and children [30]. The study found that for all the mental states' categories explored, children attribute greater mental states to children than to robots, but still attribute mental states to robots at this young age. Supporting Sommer et al. results, Kahn et al. [31] also found that children were concerned about a situation in which robots were treated in a harmful way. In their study, children (9-, 12-, and 15-year-olds) first interacted with a robot (Robovie). However, the interaction was then interrupted by placing the robot in a closet. During follow-up interviews, children expressed their perception of the robot having mental states and feelings. Children also showed strong moral convictions about how the robot should be treated. Like children, adults also anthropomorphize robots [32, 33], although they seem to do so to a lesser degree [3, 34, 35]. Such results have led to an emerging view that robots, or other embodied agents, have their own ontological category and cannot be simply compared to computers, pets, or children [36]. These studies demonstrate the value of social robots in understanding what attributes children believe may qualify an agent as having certain features of mental life. [37–40].

A large body of research in social robotics has been founded on two dimensions of mind perception: experience and agency, e.g., [41, 42]. Experience refers to one's ability to feel bodily states like hunger, pain, and pleasure. Agency refers to the extent to which one can exert self-control, plan, intend, and think. Adult humans have both of these dimensions. However, some entities have neither (e.g., a rock), whereas some have experience but very little agency (e.g., a young infant) or vice versa (e.g., a Roomba). As outlined above, social robots occupy a unique space on these two dimensions, with people thinking they have an interesting mixture of experience and agency that is different from other entities (for review see [4]).

Closely related to our study, limited work has been done to explore children's pro-social behavior towards robots [6, 7, 43]. Broadly speaking, children demonstrated pro-social behavior towards the explored robot across several studies. The extent of this pro-social behavior varied between the design and explored context of a given study, which included

either helpful behavior [7, 43] or costly and non-costly pro-social behavior [6]. These findings point to different factors that impact children's pro-social behavior towards robots, e.g., it depends on the study design and explored context. In these studies, a pro-social behavior was observed, however, it was specified either to the participants' characteristics [6] or to the explored context [7]. In Chernyak and Gary's study [6], children (5- and 7-year-old) interacted with a robot dog that was described as either moving autonomously or as remote controlled. They found that children who owned a dog were more likely to behave the autonomous robot in a pro-social way than those who did not. Martin et al. [7] explored young children's (3- to 4-year-old) helping behavior towards a robot. In their study, children witnessed a humanoid robot dropping a xylophone stick out of its reach in what seemed to be either intentionally or unintentionally. They found that children were more likely to help the robot when it dropped the stick apparently unintentionally. These results suggest that children attribute goals to a robot and are motivated to help when the robot requires help. In a different study, also exploring helpful behavior, Beran et al. found that children (5- to 16-year-old) were more likely to help a robot if they experienced a positive introduction to it, and in particular, if they were given an explicit permission to help the robot [43]. While these studies contribute to our knowledge about children's intuitions about robots as deserving pro-social behavior and kindness, they do not specifically inform whether children believe that robots have concerns surrounding distributional fairness.

There is a wealth of research demonstrating that children care deeply about fairness (for review see [44, 45]) - they believe that resources should be distributed equally between equally deserving recipients. Even young infants and toddlers show some basic expectations that resources should be divided equally and negative reactions against those who divide resources unequally [46–49]. Three-year-olds distribute equally among others [50, 51] and realize that doing so is the right thing to do [52]. From age 3 to 6, they also show negative emotional reactions to receiving less than others, becoming visibly upset when they receive fewer stickers than others [53]. As they grow older, they become increasingly likely to sacrifice their own resources in order to reduce inequality between themselves and others [54–59]. Broadly, many theories about fairness posit that infants and children attend to fairness from a young age given its importance in facilitating positive social interactions and their desire to be seen positively by others [45, 47, 60]. Even agnostic to theories about why fairness develops, this research clearly demonstrates that children's fairness concerns are ubiquitous and become more sophisticated as they mature. These concerns with fairness require children not only to track how outcomes make one individual better or worse off, but to

pay attention to the relative number of resources each person receives.

While desiring to avoid harm, be prosocial, and fair may all be thought of as morally good, there is a wealth of research suggesting that they are different kinds of concerns and may be driven by different mechanisms. Work that has examined these concerns has found that there is often a lack of correlation between helping and comforting (relating more to prosociality) and sharing with others (related more to fairness) in children [61]. Furthermore, each of these behaviors have different developmental trajectories [62], suggesting that they are not driven by the same mechanisms. Research suggests that certain behaviors (e.g. collaborating with other children and merit) more strongly prime fairness intuitions aimed at equality rather than prosociality or generosity [63]. Furthermore, comparative evidence suggests that prosociality and fairness need not be linked. Indeed, while there is extensive evidence of harm avoidance and prosocial helping in several non-human primate species, the evidence for fairness and sharing behavior is much less strong (for extensive review, see [64].) Despite that sharing exists in non-human primates, it appears to be driven by wanting to avoid harm or wanting to help others, rather than from a concern with fairness or inequality per se (for review, see [65].) Experiments that control for prosociality tend not to find that non-human primates care about inequality (e.g., [66]). Relatedly, it seems clear that it is possible that one may believe in helping agents or not harming agents without necessarily holding specific beliefs about distributional fairness. For example, although people generally think it is wrong to harm pets or that it is wrong to not *help* a pet that seems to be struggling, they might not have clear beliefs that resources should be divided equally between pets and others. Would it be unfair to give a human a treat and not a pet? What about giving a treat to a human but not a social robot? It certainly seems possible that children could believe that a social robot should not be harmed or given less, while also not believing that the social robot pays attention to distributional fairness. The current research will explore the latter question.

While many circumstances of unfairness can involve material harm (being unfair usually means that one takes resources that could go to one person and gives them to someone else), there are some fairness tasks where being unfair does not cause any material harm [56, 67–70]. These tasks in particular allow one to examine concerns with distributional fairness per se, in the absence of motivations for prosociality (for further discussion see [58, 63, 71]), which makes these tasks ideal for further probing children's intuitions about the sophistication social robots' fairness concerns. When we say "fairness" here, we refer to a concern with impartial resource distribution, what has sometimes been termed "inequity aversion" – in many cases, this will mean sharing equally between two recipients, but can mean unequal sharing when one per-

son has more merit. Here, we focus on a task developed by Shaw and Olson in which being “fair” means the extra resource is thrown in the trash, so being fair or unfair does not affect the payoff of the potentially disadvantaged agent (either way, they get nothing) [71]. The only option afforded to the distributor is to give the extra resource to “A” or throw the resource away (either way, “B” cannot get the resource). This task was specifically developed to measure children’s concern with impartial resource distribution, as separate from generosity (which is typically measured with a dictator game) and reciprocity (which is commonly studied in an ultimatum game). Indeed, when asked to distribute between A and B, the only option afforded to the distributor is to give the extra resource to “A” or throw the resource away (either way, B cannot get the resource). They developed this method to examine if children value fairness even when doing so meant being ungenerous. In this task, a child has to decide how to share between two recipients (recipient A and B) who have each received an equal number of resources. With the final resource, they can either give it to recipient A or throw the resource in the trash. Note that because of the nature of this task, being unfair here does not mean recipient B receives less - the resource will just be wasted. Still, these authors found that many children opted to throw away the resource in this case (and much more frequently than when sharing the final resource would not result in inequality). This basic pattern of results has been replicated in China, South Africa, Uganda, and the United States, with varying types of resources, with varying values, including colorful erasers, stickers, paper clips, 20 dollar bills, iPhones, and candies as full-sized Hershey bars [71–73].

Shaw and Olson’s task is ideal for exploring children’s intuitions about social robots because it focuses on a particular concern regarding whether the other agent is a being who is entitled to a certain degree of fairness or what other authors have called “respect” [45]. In particular, this task is targeted at exploring concerns with fairness *per se* in the absence of concerns for overall welfare or material harm to the agent involved. One may be wondering how this task differs from a classic task used in economics called the ultimatum game, which seems to similarly involve a decision between sharing and “destroying” resources. However, rejections in an ultimatum game need not be driven by wanting to create equal distributions between the two agents and can instead be driven by wanting to “punish” bad intentions from a distributor. This is because in an ultimatum game most decisions to reject come in response to someone who decided to be selfish. This conflates concerns with sharing equally with an aversion to negative selfish intentions from others. In addition to the Shaw and Olson task, Blake and McCaulliffe [58], developed a related task to similarly get passed these limitations of classic economic games like the dictator game and ultimatum game. As they note: “...Dictator Games assess generosity and

altruism and Ultimatum Games test strategic responses to the proposer’s intentions.” (p. 216) and are thus inappropriate for measuring something like distributional fairness *per se*. Their task and the one designed by Shaw and Olson (which we use) were designed to evaluate concerns with distributional fairness in the absence of these concerns with strategic reciprocity. It seems possible that a robot would be programmed to strategically reject in an ultimatum game, but not care at all about someone else receiving more in a case where they could not have received that resource and the person had no negative intention toward them.

Although these tasks have revealed that children think it is important to maintain fairness between other people, they do not think doing so is important when one of the “recipients” is a completely non-social agent. Indeed, even infants differentiate between social and non-social agents in their fairness expectations - they expect equal distribution to social agents, but not to non-social entities [46, 47]. Relatedly, in the throw-away paradigm used by Shaw and Olson, 6- to 8-year-old children also make different fairness decisions based on whether both recipients are social or non-social agents [71]. As reviewed above, when distributing unequal resources between two children, participants were very likely to throw away the resource rather than create inequality by giving more to one of the two recipients. However, when children were told that the other potential recipient was just a box where the experimenter would place the resources in question, then children opted to give the resource to the child.

These previous works on fairness have explored the two endpoints of person perception on the part of the recipients - the social agents were fully human and the non-social agents were “nothing” - that is, having no other recipient. Where do social robots fall on this continuum? We know that children think that one should avoid harming a social robot, but does such a robot also deserve fair treatment and, if so, what features of a social robot trigger these concerns?

2 Research Objectives

In this study, we investigated children’s evaluations and predictions about how others will respond to unfairness in order to explore the following research question: Do children believe that children and robots deserve the same fair treatment? To do so, 4- to 9-year-old children were presented with a task similar to Shaw and Olson [71], in which a distributor must decide whether to share unequally or throw a resource in the trash. They were told that the distributor ultimately decided to give the resource to one of the recipients - a child. What varied between conditions was the identity of these two recipients. In one condition, both recipients were children (Child), and in the other condition, one recipient was a child, and one was a robot (Robot-Picture Study 1, Robot-

Video Study 2). In both conditions, children were asked to respond to three measures: (i) Fairness Measure, which quantifies the extent to which children think that the allocation between two recipients is fair; (ii) Anger Measure, which quantifies how mad the non-rewarded recipient will be, and; (iii) Rectify Measure which quantifies the participants' willingness to rectify the inequality. This paradigm allowed us to explore children's reactions to the unequal sharing that occurred between these two different recipients.

The rectify measure is similar to previous work using this task, in which children are asked to make a decision about how to allocate resources. However, no previous work using this task (and very little work in developmental psychology more broadly) has probed children's evaluation of resource distributions and their feelings about it (for exceptions see [55].) Therefore, simply asking children the Fairness and Anger measures, which probe their beliefs about others' response to unequal sharing in this paradigm, is completely novel. While previous theorizing has suggested that children would believe others would be upset in such cases, it is possible that they might think: "Well, if the only other option was to throw it away, the child might not be upset." Thus, whereas previous work has primarily focused on 'fairness' as an output, we include an 'anger' measure to test whether children attribute anger from agents for not being treated as an equal recipient. Although measures like this (i.e., "how mad" with a Likert scale evaluation) have been used in previous literature with children [74], they have often not been utilized in distribution fairness tasks like the one present in these studies. Thus, these studies introduce two new aspects of examining how children think about the fairness of distributing resources to different types of agents.

Furthermore, all three measures used in this task go beyond previous work in social robotics which used tasks like the dictator and ultimatum game. These were good places to start because if children did not believe that robots deserve any kindness (dictator game) or reject unequal pay (ultimatum game), then they would be unlikely to care about the kind of distributional fairness that we are examining here. Indeed, there are several species (including dogs, monkeys, and apes) that reject unequal offers in things like ultimatum games, but never do anything that approaches the tasks we have looked at here [75]. Thus, it seems possible that children might believe that robots would reject in an ultimatum game, but not care at all about another human receiving more than them in a context where they could not have received that resource anyway.

We hypothesized that children would consider robots as less worthy of fair treatment compared to human children, particularly as children get older. We based our hypothesis on previous studies that compared other children's perceptions of children and robots, for example, their perceptions of closeness and moral concerns. In these studies, children

perceived robots as less worthy of moral concern and felt less socially close to robots than to children [2, 76]. As we previously discussed in the Introduction Section, while fairness may be thought of as similar to closeness or moral concerns, these are different concerns and therefore should be explored separately.

We also hypothesized that older children (above the age of 7 or 8) may be more likely to think that the difference between children's and robots' responses to unfairness is greater, compared to younger children. Research has demonstrated that there are considerable developments in children's fairness cognition between the ages of 4- to 9-years old [44, 52, 53, 55, 56, 58, 77]. Most significantly, during this age range, children are becoming increasingly adept at recognizing the difference between inequality that is fair and inequality that is unfair [78, 79].

Thus, we might expect that children would become more likely to differentiate their fairness evaluations of unequal sharing that disadvantages a human versus a robot. There is now over 50 years of research with children age 4 to 10 years old indicating that they can track and make decisions about how to allocate resources (for extensive reviews, see [44, 80]).

Thus, we should expect that children as young as 4 can certainly track the details of our task. This age range also seems appropriate given the work that has been done in social robotics. These social robots studies test children in varying age range, ranging between 3 years old [81], 4 to 7 years old [6, 30, 82], as well as exploring older children (5 to 8 years old [14]), and wider age range (4 to 10 years old [2]). Moreover, Sommer et al.'s findings point to the association of this age range with children's changed perceptions about robots [2].

Here, we outline our three hypotheses for the three dependent measures that we used in Experiment 1 and 2.

Hypothesis 1a. Children would think that an unequal distribution would be more unfair when it involves two children than when it involves a robot and a child.

Hypothesis 1b. Older children would regard robots as being less worthy of fair treatment than younger children.

Hypothesis 1c. The difference in perceived unfairness as a result of inequality between two children and a robot and a child will be greater for older children, compared to younger ones.

Hypothesis 2a. Children would think that a child would be angrier than a robot at an unequal distribution in which that agent receives less.

Hypothesis 2b. Older children would think robots would be less angry than younger children.

Hypothesis 2c. The difference in perceived anger from the child versus robot will be greater for older children, compared to younger ones.

Hypothesis 3a. Children would be more likely to rectify the inequality when it involves two children than when it involves a robot and a child.

Hypothesis 3b. Older children would be less willing to rectify the inequality when it involves a robot and a child than younger children.

Hypothesis 3c. The difference in their willingness to rectify inequality between two children and a robot and a child will be greater for older children than younger children.

3 Experiment 1

In the first experiment, we investigated 4- to 9-year-old children's evaluations of unequal sharing that disadvantaged either another human being or disadvantaged a robot. For specific hypotheses for this first experiment, please see research objectives.

3.1 Method

3.1.1 Participants

We recruited 139 children based in the United States: 23 4-year-olds (7 girls and 16 boys), 17 5-year-olds (8 girls and 9 boys), 25 6-year-olds (14 girls and 11 boys), 27 7-year-olds (19 girls and 8 boys), 26 8-year-olds (14 girls and 12 boys), and 21 9-year-olds (9 girls and 12 boys). The recruitment was based on a power analysis for multiple linear regression model with a view to obtaining .80 power to detect a medium effect (.25) at the standard alpha error probability (.05). Participants were recruited through a database of families who had agreed to participate in developmental research. This experiment was pre-registered on AsPredicted.¹ The study was approved by the Institutional IRB.

3.1.2 Procedure

Aiming to explore whether children believe that children and robots deserve the same fair treatment, we used a slightly modified version of Shaw and Olson's task (detailed below) [71]. These changes were made because Shaw & Olson's task asked children to distribute the resource in question whereas here we were primarily interested in how children evaluate the resource distribution as well as their predictions about how one of the recipients would react to the inequality.

Due to COVID-19, it was not possible to test children in person, so the experiment took place on Zoom. The Zoom meeting began with a short introduction of the experimenter, in which the children were told that they would be asked some questions – the kind of questions that have no right or

wrong answers. Per participant, each Zoom meeting lasted approximately 20 min. Participants were not provided with a specific context of the presented robot, such as bringing them to consider it as a robot they might interact with in the future, for example, at home or at school. The type of the presented robot was "Patricc", a non-humanoid robot, designed as an upper-torso robot's skeleton [83]. The resource used in our study was a sticker. We chose a sticker since we wanted a resource that could ostensibly be valued by both entities. We opted not to use food, for example, since the robot cannot eat food; however, children may believe that both children and robots could conceivably value stickers. Still, we acknowledge that resource value will always be a potential issue in such comparisons between distributing resources between children and robots (we return to this issue in the general discussion). Regardless of this constraint, we will be able to examine developmental differences in how children think robots and humans will respond.

Here, we present one version of our scenario that was described by the experimenter. Other versions of the scenario, not presented here, differed in factors that were counter-balanced between participants (such as the location of the winning child on the screen) and matched to participants (the gender of the recipients always matched the gender of the participant). The regular text portions are the exact wordings that were spoken to the children. The parts in italics were not spoken to the children, but describe when and where the figures were presented on the screen to the children. The Robot-Picture condition was as follows (see Fig. 5 in "Appendix B"):

Figure of a child on the left, figure of a robot on the right, and a figure of a trash can in the middle, located lower than the two other figures "Earlier today, this kid and this Robot did a really good job clearing up around the house." *A star sticker is shown in the middle, located higher than the figures of the child and the robot* "So we want to give them a sticker as a prize. But uh oh! I guess I only have one sticker to give out. Hmm."

An arrow between the sticker and the child is shown, pointing to the child "I can either give it to this kid," *The arrow is now between the sticker and the trash can, pointing to the trash* "Or I could throw it away." *A screen that was previously shown appears again with the following figures: a child, a robot, a trash can, and a sticker* "Hmm." *An arrow between the sticker and the child is shown, pointing to the child* "I think I'll give it to this kid!" *The sticker is now right above the child's figure* "See?" *An arrow pointing to the robot, while anything except for the robot and the arrow is blurred*

The Child Condition was similar, except that both recipients were merely described as children, and there were two pictures of children used (rather than a picture of a robot and a child). In both conditions, the extra sticker was given to the child recipient. See Fig. 1 for an example of the figures

¹ AsPredicted link: https://aspredicted.org/blind.php?x=YYZ_RQL.

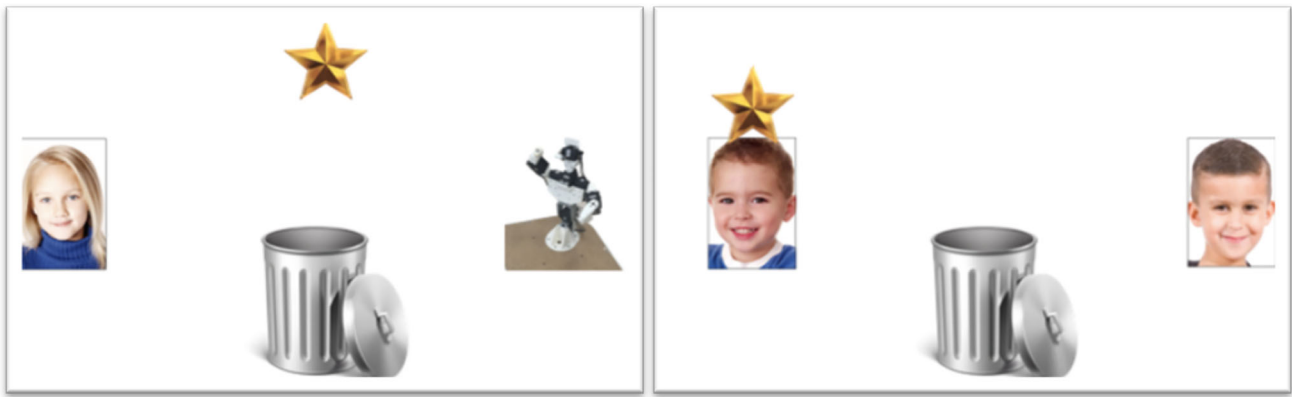


Fig. 1 Examples for the figures presented to the participants during the experiment. Left: Robot-Picture condition. Right: Child condition

presented in the Robot and child conditions from Experiment 1, and Fig. 5 in “Appendix B” for a detailed example, representing an entire conducted experiment.

After seeing one of these two conditions, children responded to three measures meant to assess their fairness intuitions. The participants were asked to evaluate how fair the allocation was (Fairness Measure), predict how mad the character (child or robot, depending on the condition) would be about receiving less (Anger Measure), and chose whether they want to rectify the inequality by taking the sticker away (Rectify Measure). For more details, please see “Measures” section. Each participant was exposed to one of the conditions. The side of the rewarded recipient, located on the right or the left, was counterbalanced between participants. We used pictures of children that were taken from a database. The pictures were normed for attractiveness and like, were counterbalanced between participants, and were gender-matched.

3.2 Measures

3.2.1 Fairness Measure

For the Fairness measure, children were asked: “How fair do you think this is? Do you think it’s fair or unfair?” The Fairness measure was a four-point Likert scale composed of four items, namely, “very unfair”, “a little unfair”, “a little fair” and “very fair”. However, the participant had to make only binary choices: the first choice was whether it was “not fair” or “fair”, followed by another binary choice of “very unfair” or “a little unfair” for the former, and “a little fair” or “very fair” for the latter. We coded this measure as an ordinal number, from 1 to 4. In these questions, we randomized the appearance of the questions: for “Do you think it’s fair? Or unfair?” - “fair” and “unfair” were randomized; and for “Would you say it’s very fair? Or a little fair?” - “very” and “little” were randomized.

3.2.2 Anger Measure

For the Anger measure, children were asked “How mad do you think the (robot or child) will be?” We used a Likert scale using four circles, from large to small. The participants had to select which circle they thought represented how much the non-recipient will be mad. We coded this measure as an ordinal number, from 1 to 4.

3.2.3 Rectify Measure

For the Rectify measure, children were asked: “Now, do you think I should take it from this kid and throw it away?” This was a binary choice of whether to take the sticker from the recipient or not: the possible answers were “Take it from the kid and throw it away!” or “Don’t take it from the kid and throw it away!”

3.3 Results

3.3.1 Fairness Measure

To test our first set of hypotheses, we performed a multilinear regression analysis (Adjusted $R^2 = .078$, $F(3,135) = 4.916$, $p = .003$) with the condition and age as predictors, controlling for gender. The regression analysis results, which are summarized in Table 1, showed a non-significant effect of gender on Fairness measure, enabling the integration of data across gender ($b = .14$, $p = .418$). Analyzing the results, we found that children believed that the distribution was more fair in the Robot-Picture Condition than the Child Condition ($b = .509$, $p = .003$), thus supporting H1a. Specifically, children rated the Robot-Picture condition as more fair than the Child condition (Robot: $M = 2.05$, $SD = 1.11$; Child: $M = 1.53$, $SD = .931$). As for age, we found an age effect on the Fairness measure, showing that older children were

Table 1 Regression analysis for Fairness

	Step 1					Step 2				
	b	SE	β	t	p	b	SE	β	t	p
Intercept	2.226	0.356		6.256	<0.001	2.676	0.455		5.884	<0.001
Condition: Robot-Picture	0.509	0.171	0.487	2.977	0.003	-0.54	0.687	-0.516	-0.786	0.434
Gender: girl	0.140	0.172	0.134	0.813	0.418	0.145	0.171	0.139	0.849	0.398
Age	-0.118	0.051	-0.189	-2.286	0.024	-0.187	0.067	-0.3	-2.771	0.006
Robot*age						0.162	0.103	0.154	1.576	0.117

Step 1: Adjusted $R^2 = .078$, $F(3,135) = 4.916$, $p = .003$. Step 2: Adjusted $R^2 = 0.088$, $F(4,134) = 4.348$, $p = .002$

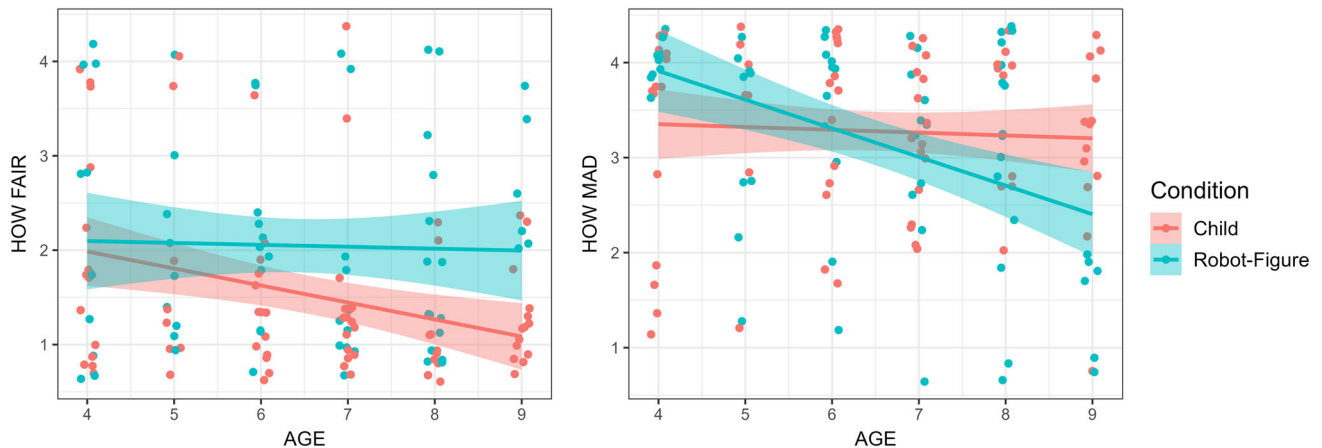


Fig. 2 Fairness and Anger score versus the participant's age, comparing the conditions of Child and Robot-Picture. Higher scores indicate more fair and more angry, respectively

more likely to believe that the distribution was unfair ($b = -.112$, $p = .03$).

To examine our second Fairness-related hypothesis, we used Pearson's correlation test in each condition. The results revealed a significant age effect on children's fairness evaluations in the Child Condition ($r(139) = -.333$, $p = .003$), as expected based on previous literature [71], but not in the Robot-Picture Condition ($r(139) = -.03$, $p = .818$), rejecting H1b. See Fig. 2 which demonstrates the difference between the conditions. It can be seen that for the Child condition, the perceived fairness decreased with age, while it remained steady for the Robot-Picture condition. Relatedly, we also examined the variance of the fairness measure among the youngest (4-years-olds) and the oldest (9-years-olds) participants, finding that these groups do not differ in their responses' variances (Levene's Test: $F(1, 43) = 2.204$, $p = .145$).

To examine our third Fairness-related hypothesis, we began with exploring the interaction effect of the condition and the age on Fairness. We conducted a separate multi-linear regression analysis to explore the interaction effect and found a non-significant effect ($b = .16$, $p = .121$). See Table 1 for both regressions results, in which "Step 1" presents the first analysis, and "Step 2" presents the regression including

the interaction term. We used Akaike Information Criterion (AIC) to compare the models. The results had shown that the model was not improved by adding the interaction term. The difference between the AIC values of each model is small (lower than two AIC units and lower than 1% change), therefore, pointing to the models' fits similarity (AIC: Step 1: 401.72, Step 2: 401.16). Thus, H1c is rejected.

3.3.2 Anger Measure

We performed similar analyses as done for the Fairness measure to test the next two sets of our hypotheses for each measure. To test the Anger measure set of hypotheses, we performed a multi-linear regression analysis (Step 1: Adjusted $R^2 = .047$, $F(3,135) = 3.261$, $p = .024$) with the condition and age as predictors, controlling for gender. Results are summarized in Table 2. Similar to the Fairness regression, we did not find a significant effect of gender on Anger Measure ($b = -.059$, $p = .716$). Therefore, we integrated the data across gender. Unlike the Fairness measure results, we found no significant difference between the Child Condition and the Robot-Picture Condition (Step 1: $b = -.117$, $p = .465$), thus rejecting H2a. In this case, higher scores indicate that the recipient who was not awarded was perceived as angrier.

Table 2 Regression analysis for Anger

	Step 1					Step 2				
	b	SE	β	t	p	b	SE	β	t	p
Intercept	4.244	0.334		12.714	< 0.001	3.485	0.418		8.342	<0.001
Condition: Robot-Picture	-0.117	0.161	-0.122	-0.731	0.466	1.651	0.631	1.71	2.616	0.01
Gender: girl	-0.059	0.162	-0.061	-0.364	0.716	-0.068	0.157	-0.07	-0.432	0.667
Age	-0.143	0.048	-0.249	-2.961	0.004	-0.03	0.062	-0.046	-0.427	0.67
Robot-Picture*age						-0.273	0.094	-0.282	-2.893	0.004

Step 1: Adjusted $R^2 = 0.047$, $F(3,135) = 3.261$, $p = .024$. Step 2: Adjusted $R^2 = .097$, $F(4,134) = 4.672$, $p = .001$

The mean scores for both conditions were similar (Robot: $M = 3.17$, $SD = 1.06$; Child: $M = 3.28$, $SD = .888$). Similar to the Fairness measure, we also found an age effect, showing that older children thought the unrewarded recipient would be less angry than younger children did ($b = -.145$, $p = .003$).

To examine our second hypothesis, we used Pearson's correlation test for each condition. The results revealed that with Robots, younger children thought that the robot would be much angrier compared to older children ($r(139) = -.467$, $p < .001$), Supporting H2b. There was no such age effect on children's expectation of how angry a child would be at receiving less ($r(139) = -.058$, $p = .622$). See Fig. 2 which demonstrates the difference between the conditions for Fairness and Anger measures. We note that the Pearson analysis is used to explore the age effect on the robot condition for each measure separately, and not to explore the possible correlation between the measures. For Anger measure, we see that the perceived Anger decreased with age for the Robot-Picture condition, while for the Child condition, it remained steady. These results can explain the rejection of H2a. The condition's null effect appears to be driven by the fact that younger children were much more likely to think that the robot would be angry than older children were. Compared to the Fairness measure, we observed the opposite results, in which we found a correlation with age in the Child Condition but not the Robot Condition.

To examine our third hypothesis, we conducted a separate multi-linear regression analysis, termed as "Step 2", to explore the interaction effect of the condition and the age on Anger (Step 2: Adjusted $R^2 = .097$, $F(4,134) = 4.672$, $p = .001$). Unlike the Fairness measure results, we found a significant interaction effect between the condition and the age ($b = -.272$, $p = .004$). Based on the AIC values, we found that adding the interaction effect improved the model's fit. The difference between the model's AIC is 6 and represents a 1.6% change (AIC: Step 1: 384.42, Step 2: 378.18). See Table 2 for the results of the regressions.

On its own, interaction analysis tells us about the existence of differences between the conditions and age, to the extent they exist. However, to fully explore our hypothesis, we used t-tests as a post-hoc analysis to compare the Anger measure of

two age groups. Specifically, we explore the edges of our age range (4- to 5- year-olds and 8- to 9-year-olds) of each condition (Child versus Robot-Picture). We found that both for older and younger children the t-test results pointed to non-significant difference between the Child and Robot-Picture condition (Older children: Robot: $M = 2.67$, $SD = 1.15$; Child: $M = 3.25$, $SD = .794$, t-test: $t(34.794) = -1.947$, $p = .06$; Younger children: Robot: $M = 3.65$, $SD = 0.813$; Child: $M = 3.32$, $SD = 1.13$, t-test: $t(38.085) = -1.1$, $p = .28$). Our results are mixed, showing that while the interaction effect was significant, the t tests showed that for both younger and older children the Anger measure was similar in both conditions. Therefore, H2c is only partially supported.

3.3.3 Rectify Measure

Lastly, we analyzed the Rectify Measure results, the third measure of our first experiment. To explore the effect of the condition on the willingness to rectify the inequality, we conducted a generalized linear model regression (GLM) using a binomial distribution. The predicted variable consisted of whether children opted to throw away the sticker or let the child keep the sticker, with the condition and age as predictors, controlling for gender. We used Area Under the Curve (AUC) and accuracy as measures to evaluate the model (AUC = 0.591, Accuracy = 58.3%). However, we found that in this case, the model demonstrated a poor discrimination (AUC < 0.7 [84]), thus rejecting H3a. See Table 3 for the regression results. However, despite the non-significant results, we noticed that the participants' responses between the conditions and the decision to rectify the inequality points to opposite trends: in the Child condition, most participants chose to throw the sticker and to rectify the inequality, while for the Robot-Picture condition, fewer participants chose to do so (Child: 58% vs. Robot-Picture: 41%). This direction of the result is in line with our hypothesis. We expected that children would be more willing to rectify the inequality between two children than a child and a robot.

Next, to explore our second Rectify-related hypothesis, we conducted a similar generalized linear mixed-effects regression, this time exploring only the Robot-Video condition (this

Table 3 Regression analysis for decision to Rectify

	Step 1				Step 2			
	Esti	SE	z	p	Esti	SE	z	p
Intercept	-0.167	0.717	-0.233	0.816	1.234	0.939	1.314	0.189
Condition: R-P ^a	0.669	0.346	1.935	0.053	-2.602	1.423	-1.829	0.067
Gender: girl	-0.015	0.348	-0.044	0.965	0.002	0.356	0.005	0.996
Age	-0.022	0.104	-0.211	0.833	-0.239	0.142	-1.686	0.092
R-P ¹ *age					0.508	0.216	2.354	0.019

Step 1: AUC = 0.591, Accuracy = 58.3%, Null deviance: 192.69 on 138 DF, Residual deviance: 188.81 on 135 DF. Step 2: AUC = 0.652, Accuracy = 64.0%, Null deviance: 192.69 on 138 DF, Residual deviance: 183.06 on 134 DF

^aRobot-Picture

way, we will be able to explore the age effect specifically in the Video-Robot condition). For this model as well, the evaluation measures pointed to insufficient model (AUC = 0.615, Accuracy = 68.3%), therefore, rejecting H3b hypothesis. Lastly, similar to previous measures' analyses, to explore our third hypothesis, we had started with conducting a regression to explore the interaction effect of the condition and age on the Rectify measure. Similar to previous regressions, the AUC measure pointed to a poor discrimination (AUC = 0.652, Accuracy = 64.0%), thus rejecting H3c. See Table 3 for the regression results.

3.4 Discussion

We found some support for our suggested hypotheses. Broadly speaking, children thought it was more unfair to share unequally when the non-recipient was a child rather than a robot. We did not find a similar main effect for Anger; however, we observed the predicted interaction with age. As children got older, they thought that a robot would be less angry about the inequality (there was no such age effect on children's predictions about the child's anger - they predicted similarly high levels of anger at all ages we tested). However, despite the observed interaction, the Anger differences between the youngest and the oldest children were not significant. Therefore, the results point to a trend, in which young children in our sample regard the two conditions as somewhat equivalent, and they increasingly differentiate between these two conditions as they get older. As for our hypothesis related to rectifying inequality, we did not find any significant differences between conditions.

These results indicate that children thought that social robots were less worthy of fair treatment than children. However, we note that these effects were quite modest compared to our expectations - children, even older children, overall believed that it was quite unfair to share unequally even if one of the recipients was a robot. Moreover, with the Anger measure, while older children thought that the robot would be less angry than younger children (interaction effect between the condition and age), even for the oldest children, we did not

observe significant anger perceptions between the Robot and the Child conditions. We hypothesized that this might have occurred because the picture of the robot, with no demonstration of the robot acting robotic, might have led children to think of the robot as being more human-like than we intended. To test this hypothesis, we conducted Experiment 2, in which we gave children a demonstration of the robot behaving more like a robot.

4 Experiment 2

In Experiment 2, we intended to further probe children's intuitions about social robots. In Experiment 1, we showed children a picture of the robot and did not give them additional information about what the robot is like. Thus, it is possible that children imagined this robot being much more agentic and human-like than we had intended. Indeed, their primary experiences with robots might be in films where robots are quite human-like [85]. If they had such an expectation, this might explain why children at all ages tested believed it was relatively unfair to share unequally between the robot and the human. To reduce the children's perceived agency of the robots, we followed previous studies findings, showing a relationship between robots' motions and their perceived animacy [86] and affect [87]. Closely related to our used approach, Castro-González et al. found that when robots moved in a mechanistic way (i.e., not smooth) they were perceived as less animate [86]. However, this finding was significant only when the robot's appearance was more human-like. The authors compared between a full-body robot (human form) and one-arm robot. Similarly, in our study, the robot resembled a human appearance, having a clear human upper-body appearance. Accordingly, in Experiment 2, we designed a video demonstration of the robot emphasizing that it was robotic. It was highly mechanistic (e.g., moving in a stilted manner) so that children understand that this robot is not particularly human-like. Similar to Experiment 1, we hypothesized that children would think that a recipient robot would be less deserving of fair treatment and also less

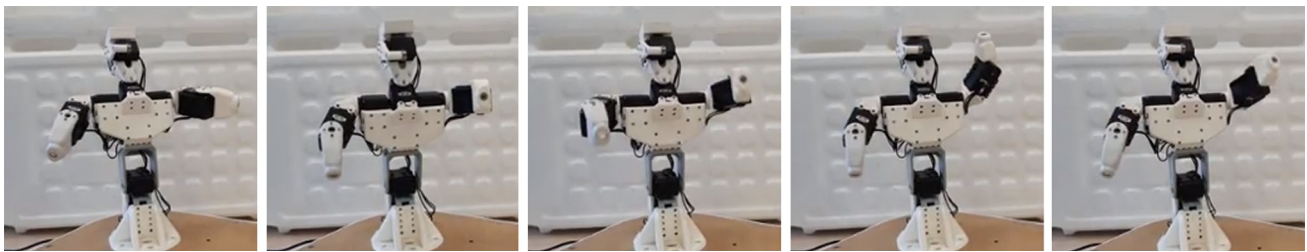


Fig. 3 Sequence of images from the video of the mechanistic robot motion

angry at unfairness, compared to a recipient child. Furthermore, based on our previous results, we hypothesized that younger children would be more likely to think that a robot recipient deserves fair treatment than older children and that younger children would believe the robot was angrier than older children. We also included the rectify measure from our previous studies, predicting that children will want to rectify the inequality more when the other recipient is a child than when it is a robot.

4.1 Method

4.1.1 Participants

We recruited 134 children based in Israel: 26 4-year-olds (11 girls and 15 boys), 28 5-year-olds (9 girls and 19 boys), 20 6-year-olds (7 girls and 13 boys), 23 7-year-olds (9 girls and 14 boys), 22 8-year-olds (6 girls and 16 boys), and 15 9-year-olds (5 girls and 10 boys). The recruitment was based on a power analysis for multiple linear regression model with a view to obtaining .80 power to detect a medium effect (.25) at the standard alpha error probability (.05). Participants were recruited through parents' groups on social networks. The study was approved by the Institutional IRB.

4.1.2 Procedure

The scenario for Experiment 2 was nearly identical to that for Experiment 1; one sticker was given to one of two recipients, and the same three dependent measures were given: Fairness Measure, Anger Measure, and Rectify Measure. However, there were two major differences between the experiments. Most importantly, children were shown a video demonstration of the robot that was meant to convey that the robot was indeed robotic. Secondly, the study used a within-subject design.

In this experiment, each participant was exposed to both conditions: Child Condition and Robot-Video Condition. The Child condition was identical to Experiment 1. The Robot-Video condition was very similar to the Robot-Picture condition from Experiment 1, except that before the trial had started, children were shown a brief video. The video con-

sisted of an 11-second video demonstration of the robot, in which the robot moved in a very mechanistic way, see Fig. 3. The robot shown in the video is the same robot shown later when describing the Robot-Video condition. The stimuli were translated into Hebrew and back translated to validate the translations. The appearance of the conditions was counterbalanced between the participants, and all other counterbalances and randomization were similar to Experiment 1 (for example, the side of the rewarded recipient and the appearance of "Fair" or "Unfair" in the relevant question).

4.1.3 Measures

The same measures were used as in Experiment 1.

4.2 Results

4.2.1 Fairness Measure

We performed similar analyses as done for Experiment 1. To test the Fairness set of hypotheses, we also used regression. However, since we conducted a within-subject study, we employed a linear mixed effect (LME) analysis to take into account the random effects in our model. The regression explored the effect of the condition and age on the Fairness Measure, controlling for gender and trial. We used the Likelihood Ratio Test (LRT) to evaluate the fixed effect significance [88] and found that the difference between the full model, which includes the condition, and the reduced model without the condition was significant ($\chi^2 = 17.092$, $p < .001$). Similar to Experiment 1, the regression analysis results, which are summarized in Table 4, showed a non-significant effect of gender on Fairness Measure, enabling integrating data across gender ($b = -.127$, $p = .415$). Similarly, the trial did not have a significant effect as well ($b = .148$, $p = .174$). Analyzing the results, we found that children believed that the distribution was more fair in the Robot-Video Condition than the Child Condition ($b = .517$, $p < .001$), thus supporting H1a. Children evaluated the Robot-Video condition as more fair than the Child condition (Robot-Video: $M = 2.37$, $SD = 1.13$; Child: $M = 1.84$, $SD = 1.02$). As for age, we did not find an age effect on Fairness measure ($b = .004$, $p = .934$).

Table 4 Regression analysis for Fairness with video demonstration

	Step 1				Step 2			
	b	SE	t	p	b	SE	t	p
Intercept	1.778	0.307	5.796	<0.001	2.549	0.369	6.905	<0.001
Condition: R-V ^a	0.517	0.113	4.581	<0.001	-0.994	0.419	-2.372	0.019
Gender: girl	-0.127	0.156	-0.818	0.415	-0.127	0.156	-0.818	0.415
Age	0.004	0.045	0.082	0.934	-0.118	0.055	-2.127	0.034
Trial	0.179	0.113	1.582	0.116	0.148	0.108	1.367	0.174
R-V ¹ *age					0.242	0.065	3.731	<0.001

Step 1: LRT between full and reduced model: $\chi^2 = 17.092$, $p < .001$. Step 2: $\chi^2 = 9.697$, $p = .002$

^aRobot-Video

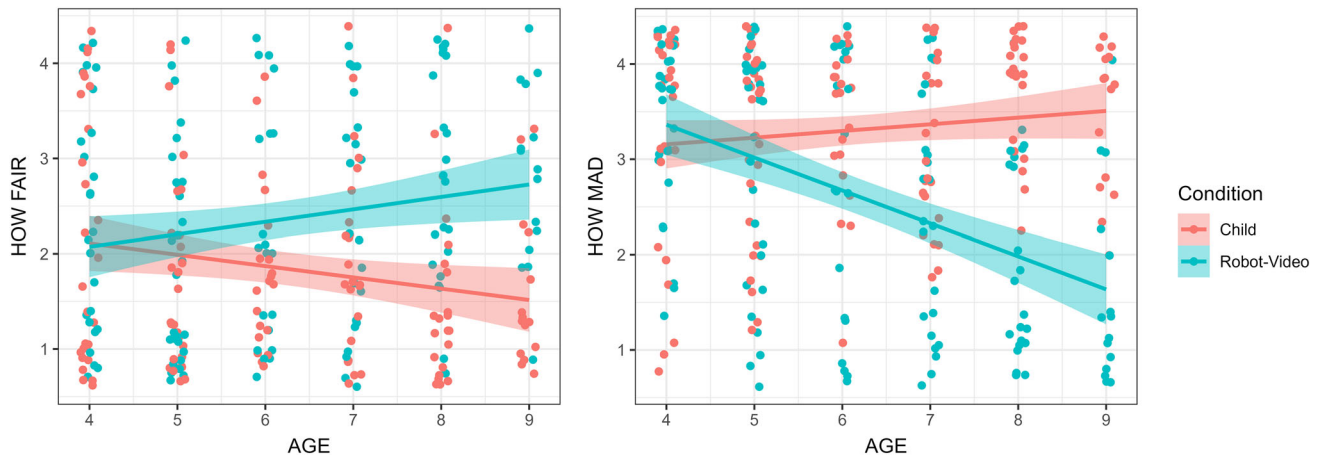


Fig. 4 Fairness score versus the participant's age, comparing the conditions of Child and Robot with video demonstration. Higher scores indicate more fair

We then used Pearson's correlation test per each condition to examine our second Fairness-related hypothesis. As expected, the results revealed a significant age effect on children's fairness evaluations in the Child Condition ($r(134) = -0.193$, $p = .025$). Older children thought that the inequality described in the experiment was more unfair compared to younger children. Unlike in Experiment 1, we also found a significant age effect in the Robot-Video Condition ($r(134) = .193$, $p = .025$), but in the opposite direction of the Child Condition. When the unrewarded recipient was a robot, older children thought that the inequality was less unfair than younger children, supporting H1b. See Fig. 4 which demonstrates the difference between the conditions. It can be seen that for the Child condition, the perceived fairness decreases with age, as opposed to the Robot-Video condition, where it increases with age. Furthermore, similar to Experiment 1, we did not find differences in the fairness responses' variances between the youngest (4-years-olds) and the oldest (9-years-olds) participants (Levene's Test: $F(1, 80) = 2.029$, $p = .158$).

Next, to examine our third Fairness-related hypothesis, we conducted a separate LME analysis, termed as "Step 2",

to explore the interaction effect of the condition and the age on Fairness. The LRT analysis showed that the difference between the full model, which includes the condition, and the reduced model without the condition was significant ($\chi^2 = 9.697$, $p = .002$). Unlike the Fairness measure results in Experiment 1, we found a significant interaction effect between the condition and the age ($b = .242$, $p < .001$). Based on the AIC values, we found that adding the interaction effect improved the model's fit. The difference between the model's AIC is 8 and represents a 1.0% change (AIC: Step 1: 814.86, Step 2: 807.18). See Table 4 for the regressions results.

Next, we used t-tests to compare the Fairness measure of the two age groups (4- to 5-year-olds and 8- to 9-year-olds) for each condition. For older children, the t-test results pointed to a significant difference between the Child and Robot-Video condition, while for the younger children the results pointed to non-significant differences (Older children: Robot-Video: $M = 2.78$, $SD = .98$; Child: $M = 1.49$, $SD = .8$, t-test: $t(36) = -6.03$, $p < .001$; Younger children: Robot-Video: $M = 2.17$, $SD = 1.15$; Child: $M = 1.98$, $SD = 1.17$, t-test: $t(53) = -.88$, $p = .38$). These results support H1c.

Table 5 Regression analysis for Anger with video demonstration

	Step 1				Step 2			
	b	SE	t	p	b	SE	t	p
Intercept	4.273	0.272	15.691	<0.001	2.986	0.348	8.571	<0.001
Condition: R-V ^a	-0.715	0.129	-5.560	<0.001	1.808	0.457	3.958	0.019
Gender: girl	0.080	0.135	0.590	0.555	0.080	0.133	0.600	0.549
Age	0.136	0.039	-3.516	0.001	0.066	0.052	1.271	0.205
Trial 2	-0.285	0.129	-2.212	0.028	-0.233	0.118	-1.980	0.050
R-V ¹ *age					-0.405	0.071	-5.717	<0.001

Step 1: LRT between full and reduced model: $\chi^2(1) = 27.076$, $p < 0.001$. Step 2: $\chi^2(2) = 26.074$, $p < 0.001$
^aRobot-Video

4.2.2 Anger Measure

Similar to Fairness analyses in Experiment 2, we began by conducting a linear mixed effect regression analysis, with the condition and age as predictors, and controlling for gender and trial (Step 1, LRT analysis results: $\chi^2 = 27.076$, $p < .001$). Since the results showed a significant effect for the trial (Step 1, LME result: $b = -0.285$, $p = .028$), in which participants were more likely to think that the unrewarded recipient is less angry in the second trial than in the first, we also analyzed the results of each trial separately. See Table 5 for the main regression results (with both trials collapsed), and see “Appendix A” for the regression of each trial separately. In what follows, we describe the main regression results. As opposed to the Anger measure in Experiment 1, in this experiment, we found a significant difference between the Child Condition and the Robot-Video Condition (Step 1: $b = -.715$, $p < .001$), thus supporting H2a. The mean scores per each condition show higher anger perceptions for the Child condition (Robot: $M = 2.59$, $SD = 1.25$; Child: $M = 3.31$, $SD = .89$). Similar to the Anger measure in Experiment 1, we found an age effect showing that older children thought that the unrewarded recipient would be less angry than younger children did ($b = -.136$, $p = .001$). In the separate analyses we conducted per each trial, we found similar results, pointing to significant effects for the condition and age (Trial 1: Condition: $b = -.477$, $p = .005$; Age: $b = -.13$, $p = .011$; Trial 2: Condition: $b = -.957$, $p < .001$, Age: $b = -.131$, $p = .025$).

We continued to examine our second hypothesis using Pearson’s correlation test. We analyzed each condition, exploring the relationship between age and Anger measure. Similar to Experiment 1, we found that for the Robot-Video condition, younger children thought that the robot would be much angrier compared to older children ($r(134) = -.462$, $p < .001$), supporting H2b, and did not find this effect for the Child condition ($r(134) = .132$, $p = .129$). Analyzing each trial, for the Robot condition, we found significant correlations in both trials (Robot-Video as a first trial: $r(65) = -.394$, $p = .001$; Robot-Video as a second trial: $r(69) = -.51$, $p < .001$). For the Child condition, we found

mixed results. Only when the Child condition was presented as the second trial we found a significant positive correlation, in which older children thought that the child would be angrier than younger children (Child as a first trial: $r(69) = -.03$, $p = .805$; Child as a second trial: $r(65) = .296$, $p = .017$). Figure 4 presents the difference between the conditions for Fairness and Anger measures. Please note that the two graphs are based on the two trials collapsed. For the Anger measure in the Robot-Video condition, we see a much sharper change with age, compared to the Child condition. For Robot-Video condition, we see a decrease in Anger perception with age, while for the Child condition, we see a mild increase with age.

Next, similar to our Fairness analysis, to examine our third hypothesis, we conducted a separate LME analysis to explore the interaction effect of the condition and the age on Anger (“Step 2”). The LRT analysis showed that the difference between the full model, which includes the condition, and the reduced model without the condition was significant ($\chi^2 = 26.074$, $p < .001$). In this case, as opposed to Step 1, the trial did not have a significant effect on Anger measure ($b = -.233$, $p = .05$). However, since this is a borderline result, we also report the regression per trial, as done in previous analyses. Similar to the Anger measure results in Experiment 1, we found a significant interaction between the condition and the age ($b = -.405$, $p < .001$). Also, similar to previous Anger results, we found that adding the interaction effect improved the model’s fit. The difference between the model’s AIC is 24.07 and represents a 3.0% change (AIC: Step 1: 812.077, Step 2: 788.004). See Table 5 for the results of the regressions. When analyzing each trial separately we also find significant interaction effect (Trial 1: $b = -.277$, $p = .006$; Trial 2: $b = -.551$, $p < .001$).

Lastly, we used t-tests to compare the Anger measure of the two age groups (4- to 5-year-olds and 8- to 9-year-olds) for each condition. We conducted this test when including both trials, and also per each trial separately. Unlike the Anger results in Experiment 1, for older children, the t-test results pointed to a significant difference between the Child and Robot-Video condition, while for the younger children the results pointed to non-significant differences (Older children:

Table 6 Regression analysis for decision to Rectify

	Step 1				Step 2			
	Esti	SE	z	p	Esti	SE	z	p
Intercept	-0.096	1.074	-0.089	0.929	1.178	1.306	0.902	0.367
Condition: R-V ^a	0.793	0.351	2.258	0.024	-1.730	1.356	-1.276	0.202
Gender: girl	-0.938	0.573	-1.637	0.102	-0.986	0.601	-1.641	0.101
Age	0.090	0.159	0.564	0.573	-0.108	0.195	-0.553	0.580
Trial 2	-0.155	0.338	-0.459	0.646	-0.205	0.348	-0.588	0.556
R-V ¹ *age					0.412	0.217	1.900	0.057
Random effects	Variance	SD			Variance	SD		
Participant	4.727	2.174			5.25	2.291		

Step 1: AUC = 0.983, Accuracy = 92.5%, logLik = -166.1, Deviance = 332.1; Step 2: AUC = 0.987, Accuracy = 92.2%, logLik = -165.9, Deviance = 331.8

^aRobot-Video

Robot-Video: $M = 1.78$, $SD = 1.0$; Child: $M = 3.59$, $SD = .644$, t -test: $t(36) = 9.571$, $p < .001$; Younger children: Robot-Video: $M = 3.15$, $SD = 1.07$; Child: $M = 3.2$, $SD = 1.03$, t -test: $t(53) = .249$, $p = .801$). These results support H2C.

For the two trials separately, we found similar results, in which only for older children there are significant differences between the conditions. (Older children: First trial: Robot-Video: $M = 2.13$, $SD = 1.19$; Child: $M = 3.32$, $SD = .716$, t -test: $t(20.95) = 3.46$, $p = 0.002$; Second trial: Robot-Video: $M = 1.55$, $SD = .8$; Child: $M = 4$, $SD = 0$, t -test: $t(21) = 14.383$, $p < .001$; Younger children: First trial: Robot-Video: $M = 3.21$, $SD = 1.03$; Child: $M = 3.35$, $SD = .977$, t -test: $t(51.975) = .482$, $p = .632$; Second trial: Robot-Video: $M = 3.08$, $SD = 1.13$; Child: $M = 3.07$, $SD = 1.09$, t -test: $t(51.336) = -.018$, $p = .986$).

4.2.3 Rectify Measure

Lastly, we analyzed the Rectify Measure results. To explore the effect of the condition on the willingness to rectify the inequality, we used generalized linear mixed-effects analysis, using a binomial distribution. The predicted variable in the regression consisted of whether to throw or let the child keep the sticker, with the condition and age predictors, and controlling for gender and trial. Unlike Experiment 1, in this experiment, the evaluation measures pointed to a sufficient model (AUC = 0.983, Accuracy = 92.5%). We observed a significant effect of the condition on the Rectify Measure (coefficient = 0.793, $p = 0.024$), in which children were more likely to keep the sticker in the Robot-Video condition, i.e. not rectifying the inequality, thus, supporting H3a. See Table 6 for the regression result.

Next, similar to Experiment 1, we conducted a generalized linear mixed-effects regression while exploring only the Robot-Video condition. In this model, the evaluation measures pointed to an insufficient model (AUC = 0.643, Accuracy = 61.9%), therefore, rejecting H3b. Lastly, to inves-

tigate our third Rectify-related hypothesis, we explored the interaction effect of the condition and age on the Rectify measure. In this case, while the model itself was sufficient (AUC = 0.987, Accuracy = 92.2%), the results pointed to insignificant interaction effect (coefficient = 0.412, $p = 0.057$), thus rejecting H3c. See Table 6 for the regression result.

4.3 Discussion

In Experiment 2, we replicated key findings from Experiment 1 and found support for some hypotheses that only received mixed support in Experiment 1. Summary of supported and rejected hypotheses in both experiments is presented in Table 9 in "Appendix C". Replicating Experiment 1, we found that children thought it was more fair to share unequally when one of the recipients was a robot rather than when both recipients were children. Additionally, we found the predicted change in children's fairness evaluations as they got older. Younger children thought that it was equally unfair to give the resource to one child over another child as it was to give the resource to one child over a robot. As children got older, they increasingly thought it was much more unfair to share unequally when the other recipient was a child rather than a robot. Indeed, this interaction of age and condition shows that while younger children do not necessarily differentiate between how children and social robots should be treated in this task, older children think they should be treated differently. This is similar to the attribution of mental states to robots that have been found for 5- to 6-years-olds in previous work [30] - while they significantly recognized the robot as a distinct entity, children of this age expressed similar behaviors toward a robot and a child. Moreover, the age dependency shown in our results may reflect the over-generalization of fairness concerns [89] as well as younger children's tendency to anthropomorphize certain inanimate objects [34, 90, 91]. Our results reveal striking developments in children's evaluations and beliefs about how robots should be treated.

As for Anger, we replicated our results from Experiment 1, again finding a significant interaction between Anger and age such that, with age, children increasingly believed that the robot would be less angry at inequality. However, only in Experiment 2 we also found that younger children thought that a child and robot would be upset in a similar way about receiving less, while older children thought that a robot would be much less upset than a child. Additionally, in this experiment, we also observed an overall main effect of condition such that children thought that a child would be more upset than a robot as receiving less (though this main effect was qualified by the interaction we just described). These results suggest that young children believe that robots will be less upset about inequality than other children and, specifically, that they do so increasingly with age.

Finally, unlike in Experiment 1, we found that children differentiated between the child and robot recipients in their tendency to rectify inequality. We had predicted that children would be more likely to rectify the inequality (taking the sticker back from the child recipient) when the other disadvantaged recipient was a robot rather than a human, and we found support for this prediction. However, unlike the other measures (Fairness, Anger), we did not find that age had an effect on children's likelihood to rectify the inequality.

5 General Discussion

In two studies, we demonstrated first, that children think it is more unfair to share unequally when both recipients are children as compared to when one of the recipients is a child, and the other is a robot; and second, that children evaluate human children as more likely to be mad than robots when they receive fewer resources from a distributor. Both of these tendencies appeared to become stronger with age, with our oldest children most clearly believing it is more unfair to share unequally between two children as compared to a child and a robot. We do not believe that this is because younger children could not follow our task or thought differently about fairness—4-year-olds did not show more variance than the older children on our fairness measure in both experiments, and all age groups generally regarded unequal sharing between human recipients as unfair (see explanation later in our general discussion)—but instead suggest that younger children specifically thought differently about inequality involving a robot recipient. Supporting Castro-González et al. results' [86], we further demonstrated that children's expectations appear to be influenced by how robotic the robot seems to children. Especially for the older children in our sample, they most strongly differentiated between robots and child recipients when they were shown a video in which the robot moved very mechanically (Experiment 2). Taken together, these results reveal that young children appear to treat robots and human beings as some-

what equivalent on this task, but as they get older, they clearly predict that robots (at least when they are presented as highly robotic as in Experiment 2) will be much less angry about inequality and evaluate creating such inequality as less unfair.

These studies help to sketch a more detailed picture of children's early fairness intuitions and how they think about non-human agents as deserving or not deserving fair treatment. Previous research in this area has found that children from early in life (even infants) differentiate between unequal sharing that occurs between agents and non-agents [46–48]. However, the way they think about and evaluate unequal sharing between different agents (who might differ in their ability to experience certain kinds of emotions) has been less clear. Furthermore, although we know that children do get upset when they receive less than others [53, 55], there has been much less work that we are aware of that has examined children's predictions about how others will feel about inequality (for an exception see [92]) and no work that has explicitly examined how children believe that non-human agents feel about inequality. These studies thus make two clear advances. They demonstrate that children do predict that others will be mad at inequality, even in cases where the only alternative was to throw the resource away. Further, children, especially as they get older, modulate their predictions about others' anger based on the identity of the disadvantaged agent (i.e. whether they are a robot or not). Future work should further probe how children's beliefs and predictions about others' reactions to inequality drive the way they interact with and share with others.

Furthermore, these studies make an important methodological point for those interested in measuring children's intuitions about robots. It has long been known that children's tendency to anthropomorphize a robot is driven by how human-like the robot seems [3, 93, 94]. Based on these data, one might have assumed that when children hear the word robot and see a picture like the one used in our Experiment 1, they would infer a very robotic and non-humanlike entity. However, our results suggest that children instead seemed to infer a more human-like robot (especially when they were young). It was only when we gave them positive evidence of the robot being robotic that they thought of this robot as being less human.

Most of children's exposure to robots is likely from TV and film, in which the robots are highly human-like [85]. Hence, it is not surprising that their initial perception of a robot is primed to anthropomorphic robots, with human-like behaviors [95]. We assume that these perceptions affected their responses. Hence, if children are presented with only a picture of a robot, as was the case in Experiment 1, they may imbue the robot with unintended human-like qualities that could induce a higher level of moral attribution towards it, resulting in smaller differences between how they believe a child and a robot should be treated. Importantly, our studies

demonstrated that their initial expectations can be overcome by showing children that the robot is actually quite robotic. Indeed, we demonstrate that even a short video, as was used in Experiment 2, that conveys the mechanistic characteristic of the robot (e.g., it did not speak or interact and moves in a jerky “robotic” way) was able to change children’s beliefs about the robot, resulting in larger and significant differences between children’s fairness evaluations of unequal sharing that disadvantaged a child versus a robot. These data will be useful for subsequent work on children’s intuitions about robots because they suggest that children may need some actual exposure to the robot (at least in a video) in order to actually probe children’s intuitions about the kinds of robots that exist in the real world. For example, for studies in the HRI field [96], as well as outside of it, such as in Psychology, using a graphical demonstration of robots is common. For these studies, clearly providing a demonstration of the extent to which a robot is human-like would be useful, as we observed that children have prior perceptions regarding this topic.

While these studies further our understanding of children’s concepts of fairness and their intuitions about robots, they are not without their limitations. First of all, the studies were conducted online, due to the COVID-19 pandemic. These circumstances raise two separate issues. First, while we intended to explore our hypotheses using a physical robot that would have enabled an interaction, we could only explore our hypotheses using an image and a video, a methodology used in prior children-robots studies [2, 30, 97]. It is possible that interacting with a real robot might influence children’s fairness evaluations. For example, seeing robots respond to external stimuli, such as responding to pain, may drive up even older children’s fairness concerns. Alternatively, a very robotic robot may drive down children’s fairness concerns, possibly even in young children. Given that we saw some effect of children’s observations from video affecting fairness measures, this seems quite likely. Indeed, the difference between our two studies demonstrates to some degree that children are sensitive to how the robot behaves even when they encounter a photo or video of a robot virtually. Thus, using virtual representations of robots for convenience can provide some methodological insights for future online studies. Our study results highlight the importance of understanding what the children’s perceptions about a presented robot are, and according to the research purpose, clarifying the extent to which the explored robot is human-like. Our studies demonstrate that merely showing children a robot is probably insufficient, and research procedures should identify some basic features of the explored robot to shift children’s intuitions about the robot [97]. If we were to have children interact with a real-life robot (who acted robotically), we believe this too might shift them to think of the robot as less human. However, we hypothesize that the general trend observed will remain, leading to possibly stronger

results (such as a stronger effect of age), but not to a different pattern entirely. Second, previous research in our lab and other labs has used Zoom or other video-conference services, as a platform to conduct studies in other contexts, such as conducting interviews and focus groups. These studies found that the experience of video-conferencing platforms was overall positive and comparable to in-person settings [98–100].

Next, we were only able to test children’s intuitions about a few specific robots and human beings, but of course, there are a number of entities that are less human-like than the somewhat humanoid-shaped social robots that appeared in our studies. Pets, non-humanoid robots, and toys are three entities that likely would engender their own response from children. Because we did not include a highly non-social control (e.g. a toy car that is not a robot at all) we do not know how much children’s responses were shaped by their concepts of robots in particular versus their intuitions about anything non-human. We noted above that younger children may have thought robots would be mad because they anthropomorphize robots, but it could be that younger children would anthropomorphize any entity. It will be highly informative for future research in this area to include completely non-social controls in order to assess if young children at least have the intuition that these entities will not experience anger or deserve fair treatment. If, instead, young children respond the same way for completely non-social controls, this would suggest that younger children’s responses are not specific to robots and instead part of some broader tendency to attribute anger to inanimate things.

Another potential limitation of the study is that it did not include an example-based explanation to help the participants understand the concepts and validate their understanding prior to the formal experiments. We did this because our task was relatively straight forward and so we believed that the bit of scaffolding we did provide children was sufficient. Furthermore, if we want to assess children’s spontaneous judgments, then it can be better to do so without too much formal training (as would be the case in their normal lives). Of course, we acknowledge that there are cases where such a lack of training can potentially be a problem, particularly when interpreting the behavior of younger children. For example, one may find that younger children respond differently from older children simply because they have difficulty following the task at hand. While this is always a possibility, there are features of our results that make such an interpretation unlikely to account for our results. If young children were merely having difficulty following our task, then they should have responded more randomly than the older children (primarily responding at the midpoints of our scales). However, that is not at all what we find, we instead find that younger children quite strongly think that both robots and children will be very mad. Furthermore, younger and older children generally respond similarly to the question of whether or not

a child would be mad about inequality in our two studies. Such systematic responding cannot occur with children simply misunderstanding the task and so we do not think this should undermine the interpretation of our results here.

Finally, although we tested these intuitions in two cultures (the United States and Israel), the two cultures we tested are highly similar on many dimensions, including the extent to which they are rich and industrialized. In some ways, this was a natural starting point for such an investigation, given that children in these societies will be the ones most likely to interact with social robots in the near future. However, these studies obviously cannot tell us broader things about how children in other countries, i.e., less industrialized societies, with less exposure to robots, think about social robots. Indeed, we know that culture has pervasive effects on people's judgments, which has created a recent push to test people beyond WEIRD societies (western, educated, industrialized, rich and democratic) [101]. It would certainly seem that people's intuitions about robots will likely be influenced both by their cultural experience as well as just their exposure to social robots. Indeed, we know that children respond differently to robots, even in industrialized societies, based on their experiences with them [82, 102]. For children who have very limited or no interaction with social robots, it is unclear how they might evaluate such robots. One possibility is that if they had no exposure at all and were shown the non-biological robotic video that participants were shown in Experiment 2, they would not believe that the robot deserves fair treatment. However, there is some work suggesting that, at least in industrialized societies, more exposure to social robots leads to children anthropomorphizing them less and so it is possible that children in non-industrialized societies would actually anthropomorphize the robots more. We hope that this work prompts more cross-cultural investigations on children's intuitions about robots [82]. Future work can address these issues by introducing more background information on each participant, such as prior experience with robots, e.g. in media exposure, interaction with technology and technological toys. Furthermore, including a larger variation of robots on the anthropomorphic-axis will pinpoint the morphological aspects that trigger the effects reported here.

As we noted at the outset, social robots are not merely the future, they are already here. Not only are these social robots important tools that can help educate our children, but probing children's intuitions about these social robots can educate researchers interested in people's basic social cognition. Our results reveal important developments in children's intuitions about fairness and social robots, and thus opens up the possibility for countless future research: investigating the features that drive one to think that others will be upset about inequality, as well as broader questions about what kinds of rights people believe that robots are entitled to.

Acknowledgements This work was supported by the Jacobs Foundation, grant number 2019131700, and by the Shulamit Aloni Scholarship by the Israeli Ministry of Science and Technology, grant number 314575.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability. The data that support the findings of this study are openly available in OSF: https://osf.io/awqmy/?view_only=2b019522443f429ca698145f954150a8

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A. Experiment 2, Anger Measure: Separated Trials Regressions

Table 7 Regression analysis for Anger, study 2, trial 1

	Step 1					Step 2				
	b	SE	β	t	p	b	SE	β	t	p
Intercept	4.030	0.353	0.109	11.407	<0.001	3.238	0.447		7.242	<0.001
Condition: Robot-Video	-0.477	0.169	-0.464	-2.822	0.006	1.248	0.642	1.214	1.944	0.054
Gender: girl	0.340	0.177	0.331	1.922	0.057	0.352	0.173	0.342	2.038	0.044
Age	-0.130	0.051	-2.561	-0.212	0.012	-0.006	0.067	-0.011	-0.097	0.923
Robot-Video*age						-0.277	0.100	-0.270	-2.781	0.006

Step 1: Adjusted $R^2 = 0.01$, $F(3,130) = 5.912$, $p = 0.001$. Step 2: Adjusted $R^2 = 0.144$, $F(4,129) = 6.597$, $p = < 0.001$

Table 8 Regression analysis for Anger, study 2, trial 2

	Step 1					Step 2				
	b	SE	β	t	p	b	SE	β	t	p
Intercept	4.172	0.390	0.454	10.706	< 0.001	2.319	0.504		4.599	< 0.001
condition: Robot-Video	-0.957	0.192	-0.781	-4.977	< 0.001	2.471	0.683	2.018	3.616	< 0.001
Gender: girl	-0.181	0.201	-0.148	-0.903	0.368	-0.205	0.184	-0.167	-1.114	0.267
Age	-0.131	0.058	-0.179	-2.264	0.025	0.174	0.079	0.237	2.200	0.030
Robot-Video*age						-0.551	0.106	-0.450	-5.190	< 0.001

Step 1: Adjusted $R^2 = 0.181$, $F(3,130) = 10.78$, $p = < 0.001$. Step 2: Adjusted $R^2 = 0.317$, $F(4,129) = 16.43$, $p = < 0.001$

Appendix B. A Full Example of the Robot-Picture Condition

Appendix C. Summary of Results

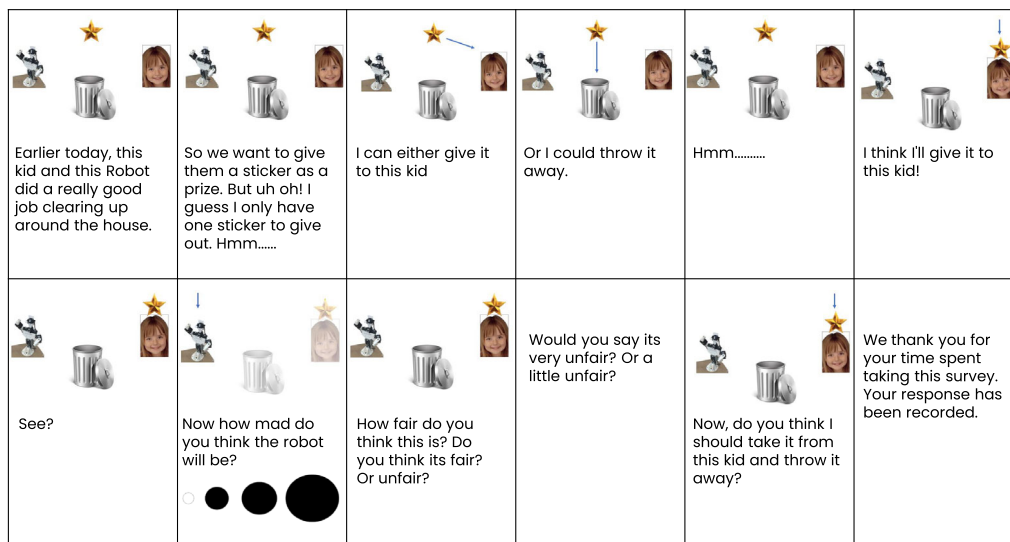


Fig. 5 An example of a Robot-Picture condition as was conducted in Experiment 1. In this case, the child was a girl, positioned on the right side

Table 9 supported and rejected hypotheses in Experiment 1 and Experiment 2

Measure	Hypothesis	Experiment 1	Experiment 2
Fairness	H1a	Supported	Supported
	H1b	Rejected	Supported
	H1c	Rejected	Supported
Anger	H2a	Rejected	Supported
	H2b	Supported	Supported
	H2c	Partially supported	Supported
Rectify	H3a	Rejected	Supported
	H3b	Rejected	Rejected
	H3c	Rejected	Rejected

Hypotheses: Ha: comparing children's perceptions of child-child versus child-robot context; Hb: age related, hypothesizing older children will consider robots as less worth of the explored measure. Hc: age related and comparing perceptions about child-child versus child-robot context. See Sect. 2 for the detailed hypotheses

References

- Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. *Sci Robot* 3(21):eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>
- Sommer K et al (2019) Children's perceptions of the moral worth of live agents, robots, and inanimate objects. *J Exp Child Psychol* 187:104656. <https://doi.org/10.1016/j.jecp.2019.06.009>
- Tung F-W (2016) Child perception of humanoid robot appearance and behavior. *Int J Hum Comput Interact* 32(6):493–502. <https://doi.org/10.1080/10447318.2016.1172808>
- Bigman YE, Waytz A, Alterovitz R, Gray K (2019) Holding robots responsible: the elements of machine morality. *Trends Cognit Sci* 23(5):365–368
- Van Straten CL, Peter J, Kühne R, De Jong C, Barco A (2018) Technological and interpersonal trust in child-robot interaction: an exploratory study, pp 253–259
- Chernyak N, Gary HE (2016) Children's cognitive and behavioral reactions to an autonomous versus controlled social robot dog. *Early Educ Dev* 27(8):1175–1189
- Martin DU et al (2020) Investigating the nature of children's altruism using a social humanoid robot. *Comput Hum Behav* 104:106149
- Zhou J, Verma S, Mittal M, Chen F (2021) Understanding relations between perception of fairness and trust in algorithmic decision making. *IEEE*, pp 1–5
- Starke C, Baleis J, Keller B, Marcinkowski F (2022) Fairness perceptions of algorithmic decision-making: a systematic review of the empirical literature. *Big Data Soc* 9(2):20539517221115188
- Saunderson S, Nejat G (2019) How robots influence humans: a survey of nonverbal communication in social human-robot interaction. *Int J Soc Robot* 11(4):575–608. <https://doi.org/10.1007/s12369-019-00523-0>
- Kennedy J, Baxter P, Belpaeme T (2017) The impact of robot tutor nonverbal social behavior on child learning. *Front ICT*. <https://doi.org/10.3389/fict.2017.00006>
- Breazeal C, Scassellati B (1999) How to build robots that make friends and influence people, vol 2, pp 858–863
- Gordon G, Breazeal C, Engel S (2015) Can children catch curiosity from a social robot?, vol 15. ACM, New York, pp 91–98
- Short E et al (2014) How to train your DragonBot: socially assistive robots for teaching children about nutrition through play. *IEEE*, Edinburgh, pp 924–929
- Fasola J, Mataric M (2013) A socially assistive robot exercise coach for the elderly. *J Hum Robot Interact* 2(2):3–32. <https://doi.org/10.5898/JHRI.2.2.Fasola>
- Gordon G et al (2016) Affective personalization of a social robot tutor for children's second language skills. <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11759>
- Vogt P et al (2019) Second Language Tutoring Using Social Robots: A Large-Scale Study, 497–505. ISSN 2167–2148:2167–2121
- Amazon. Astro (2023). <https://www.amazon.com/Introducing-Amazon-Astro/dp/B078NSDFSB>
- LuxAI. QTrobot (2023). <https://luxai.com/>
- Embodied. Moxie (2023). <https://embodied.com/>
- Mubin O, Stevens CJ, Shahid S, Mahmud AA, Dong J-J (2013) A review of the applicability of robots in education. *J Technol Educ Learn* 1(209–0015):13. <https://doi.org/10.2316/Journal.209.2013.1.209-0015>
- Kory-Westlund JM, Breazeal C (2019) A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front Robot AI*. <https://doi.org/10.3389/frobt.2019.00081>
- Gordon G, Breazeal C (2015) Bayesian active learning-based robot tutor for children's word-reading skills
- Shiomi M, Kanda T, Howley I, Hayashi K, Hagita N (2015) Can a social robot stimulate science curiosity in classrooms? *Int J Soc Robot* 7(5):641–652. <https://doi.org/10.1007/s12369-015-0303-1>
- Ali S, Moroso T, Breazeal C (2019) Can children learn creativity from a social robot?, vol 19. ACM, New York, pp 359–368
- Park HW, Rosenberg-Kima R, Rosenberg M, Gordon G, Breazeal C (2017) Growing growth mindset with a social robot peer, vol 17. ACM, New York, pp 137–145
- Sisman B, Gunay D, Kucuk S (2019) Development and validation of an educational robot attitude scale (ERAS) for secondary school students. *Interact Learn Environ* 27(3):377–388. <https://doi.org/10.1080/10494820.2018.1474234>
- Tahir Y, Rasheed U, Dauwels S, Dauwels J (2014) Perception of humanoid social mediator in two-person dialogs, HRI '14. ACM, New York, pp 300–301. <https://doi.org/10.1145/2559636.2559831>
- Rosenberg-Kima RB, Koren Y, Gordon G (2020) Robot-supported collaborative learning (RSCL): social robots as teaching assistants for higher education small group facilitation. *Front Robot AI*. <https://doi.org/10.3389/frobt.2019.00148>
- Di Dio C et al (2019) It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. *Int J Soc Robot*. <https://doi.org/10.1007/s12369-019-00528-9>
- Kahn PH Jr et al (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>
32. Hayes B, Ullman D, Alexander E, Bank C, Scassellati B (2014) People help robots who help others, not robots who help themselves, pp 255–260
 33. Short E, Hart J, Vu M, Scassellati B (2010) No fair. An interaction with a cheating robot, pp 219–226
 34. Severson RL, Lemm KM (2016) Kids see human too: adapting an individual differences measure of anthropomorphism for a child sample. *J Cognit Dev* 17(1):122–141
 35. van Straten CL, Peter J, Kühne R (2020) 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>
 36. Kahn PH et al (2011) The new ontological category hypothesis in human–robot interaction, pp 159–160
 37. Geraci A, Simion F (2021) Evaluation of prosocial actions performed by dynamic shapes at 17 months of age. *Eur J Dev Psychol* 8:1–25
 38. Hamlin JK, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. *Nature* 450(7169):557–559
 39. Johnson SC et al (2010) At the intersection of social and cognitive development: internal working models of attachment in infancy. *Cognit Sci* 34(5):807–825
 40. Kanakogi Y, Okumura Y, Inoue Y, Kitazaki M, Itakura S (2013) Rudimentary sympathy in preverbal infants: preference for others in distress. *PLoS ONE* 8(6):e65292
 41. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. *Science* 315(5812):619–619
 42. Gray K, Young L, Waytz A (2012) Mind perception is the essence of morality. *Psychol Inq* 23(2):101–124
 43. Beran TN, Ramirez-Serrano A, Kuzyk R, Nugent S, Fior M (2011) Would children help a robot in need? *Int J Soc Robot* 3(1):83–93
 44. Hook JG, Cook TD (1979) Equity theory and the cognitive ability of children. *Psychol Bull* 86(3):429
 45. Engelmann JM, Tomasello M (2019) Children’s sense of fairness as equal respect. *Trends Cogn Sci* 23(6):454–463
 46. Geraci A, Surian L (2011) The developmental roots of fairness: infants’ reactions to equal and unequal distributions of resources. *Dev Sci* 14(5):1012–1020
 47. Schmidt MF, Sommerville JA (2011) Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE* 6(10):e23223
 48. Sloane S, Baillargeon R, Premack D (2012) Do infants have a sense of fairness? *Psychol Sci* 23(2):196–204
 49. Geraci A, Nuovo SD (2018) Toddlers’ spontaneous reactions to inequality in distributive actions. *J Child Adolesc Behav* 6(379):2
 50. Olson KR, Spelke ES (2008) Foundations of cooperation in young children. *Cognition* 108(1):222–231
 51. Paulus M, Christner N, Wörle M (2020) The normative status of friendship: do young children enforce sharing with friends and appreciate reasonable partiality? *J Exp Child Psychol* 194:104826
 52. Smith CE, Blake PR, Harris PL (2013) I should but I won’t: why young children endorse norms of fair sharing but do not follow them. *PLoS ONE* 8(3):e59510
 53. LoBue V, Nishida T, Chiong C, DeLoache JS, Haidt J (2011) When getting something good is bad: even three year olds react to inequality. *Soc Dev* 20(1):154–170
 54. Blake PR, Rand DG (2010) Currency value moderates equity preference among young children. *Evol Hum Behav* 31(3):210–218
 55. Kogut T (2012) Knowing what I should, doing what I want: from selfishness to inequity aversion in young children’s sharing behavior. *J Econ Psychol* 33(1):226–236
 56. Shaw A, Choshen-Hillel S, Caruso EM (2016) The development of inequity aversion: understanding when (and why) people give others the bigger piece of the pie. *Psychol Sci* 27(10):1352–1359
 57. Sigelman CK (2012) Rich man, poor man: developmental differences in attributions and perceptions. *J Exp Child Psychol* 113(3):415–429
 58. Blake PR, McAuliffe K (2011) “I had so much it didn’t seem fair”: eight-year-olds reject two forms of inequity. *Cognition* 120(2):215–224
 59. Moore C (2009) Fairness in children’s resource allocation depends on the recipient. *Psychol Sci* 20(8):944–948
 60. Shaw A et al (2014) Children develop a veil of fairness. *J Exp Psychol Gen* 143(1):363
 61. Dunfield K, Kuhlmeier VA, O’Connell L, Kelley E (2011) Examining the diversity of prosocial behavior: helping, sharing, and comforting in infancy. *Infancy* 16(3):227–247
 62. Paulus M (2014) The emergence of prosocial behavior: why do infants and toddlers help, comfort, and share? *Child Dev Perspect* 8(2):77–81
 63. Corbit J, McAuliffe K, Callaghan TC, Blake PR, Warneken F (2017) Children’s collaboration induces fairness rather than generosity. *Cognition* 168:344–356
 64. Warneken F, Tomasello M (2015) The developmental and evolutionary origins of human helping. *The Oxford handbook of prosocial behavior*. Oxford Library of Psychology
 65. Sheskin M, Santos L (2012) The evolution of morality: which aspects of human moral concerns are shared with nonhuman primates. *Oxford Handb Comp Evoluti Psychol* 13:434–449
 66. McAuliffe K et al (2015) Capuchin monkeys, *Cebus apella*, show no evidence for inequity aversion in a costly choice task. *Anim Behav* 103:65–74
 67. Blake PR et al (2015) The ontogeny of fairness in seven societies. *Nature* 528(7581):258–261
 68. Corbit J, MacDougall H, Hartlin S, Moore C (2022) The development of intergroup cooperation: children show impartial fairness and biased care. *Front Psychol* 13:825987
 69. Gordon-Hecker T, Shaw A, Choshen-Hillel S (2022) One for me, two for you: agency increases children’s satisfaction with disadvantageous inequity. *J Exp Soc Psychol* 100:104286
 70. Kenward B, Dahl M (2011) Preschoolers distribute scarce resources according to the moral valence of recipients’ previous actions. *Dev Psychol* 47(4):1054
 71. Shaw A, Olson KR (2012) Children discard a resource to avoid inequity. *J Exp Psychol Gen* 141(2):382
 72. Choshen-Hillel S, Lin Z, Shaw A (2020) Children weigh equity and efficiency in making allocation decisions: evidence from the US, Israel, and China. *J Econ Behav Org* 179:702–714
 73. Paulus M (2015) Children’s inequity aversion depends on culture: a cross-cultural comparison. *J Exp Child Psychol* 132:240–246
 74. Sachs ME, Kaplan J, Habibi A (2019) Echoing the emotions of others: empathy is related to how adults and children map emotion onto the body. *Cogn Emot* 33(8):1639–1654
 75. Brosnan SF, de Waal FB (2014) Evolution of responses to (un) fairness. *Science* 346(6207):1251776
 76. Shahid S, Krahmer E, Swerts M (2014) Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Comput Hum Behav* 40:86–100. <https://doi.org/10.1016/j.chb.2014.07.043>
 77. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454(7208):1079–1083. <https://doi.org/10.1038/nature07155>
 78. Schmidt MF, Svetlova M, Johe J, Tomasello M (2016) Children’s developing understanding of legitimate reasons for allocating resources unequally. *Cogn Dev* 37:42–52
 79. Shaw A, Olson K (2014) Fairness as partiality aversion: the development of procedural justice. *J Exp Child Psychol* 119:40–53
 80. McAuliffe K, Blake PR, Steinbeis N, Warneken F (2017) The developmental foundations of human fairness. *Nat Hum Behav* 1(2):0042

81. Martin DU et al (2020) Young children's indiscriminate helping behavior toward a humanoid robot. *Front Psychol* 11:239. <https://doi.org/10.3389/fpsyg.2020.00239>
82. Bernstein D, Crowley K (2008) Searching for signs of intelligent life: an investigation of young children's beliefs about robot intelligence. *J Learn Sci* 17(2):225–247
83. Gvirsman O, Koren Y, Norman T, Gordon G (2020) Patricc: a platform for triadic interaction with changeable characters, vol 20. Association for Computing Machinery, Cambridge, pp 399–407
84. Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, vol 398. Wiley
85. Casiddu N, Porfirione C, Burlando F, Vacanti A, Zallio M, Raymundo Ibañez C, Hernandez JH (2021) A century of humanoid robotics in cinema: a design-driven review. In: Zallio M, Raymundo Ibañez C, Hernandez JH (eds) *Advances in human factors in robots, unmanned systems and cybersecurity*. Lecture notes in networks and systems. Springer, Cham, pp 103–109
86. Castro-Gonzalez AH, Scassellati B (2016) Effects of form and motion on judgments of social robots animacy, likability, trustworthiness and unpleasantness. *Int J Hum Comput Stud* 90:27–38
87. Saerbeck M, Bartneck C (2010) Perception of affect elicited by robot motion. *IEEE*, pp 53–60
88. Meteyard L, Davies RA (2020) Best practice guidance for linear mixed-effects models in psychological science. *J Memory Lang* 112:104092
89. Schmidt SMP, Ralph DL (2016) The flipped classroom: a twist on teaching. *Contemp Issues Educ Res* 9(1):1–6. <https://doi.org/10.19030/cier.v9i1.9544>
90. Gjersoe NL, Hall EL, Hood B (2015) Children attribute mental lives to toys when they are emotionally attached to them. *Cogn Dev* 34:28–38
91. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev* 114(4):864
92. Paulus M, Moore C (2015) Preschool children's anticipation of recipients' emotions affects their resource allocation. *Soc Dev* 24(4):852–867
93. Blancas M et al (2017) Analyzing children's expectations from robotic companions in educational settings, pp 749–755
94. Woods S, Dautenhahn K, Schulz J (2004) The design space of robots: investigating children's views, pp 47–52
95. Sandoval EB, Mubin O, Obaid M (2014) Human robot interaction and fiction: a contradiction. Springer, pp 54–63
96. Law T, Chita-Tegmark M, Scheutz M (2021) The interplay between emotional intelligence, trust, and gender in human–robot interaction: a vignette-based study. *Int J Soc Robot* 13(2):297–309
97. Randall N, Sabanovic S (2023) A picture might be worth a thousand words, but it's not always enough to evaluate robots, vol 23. Association for Computing Machinery, New York, pp 437–445. <https://doi.org/10.1145/3568162.3576970>
98. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M (2019) Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. *Int J Qual Methods* 18:1609406919874596
99. Falter M M et al (2022) Making room for Zoom in focus group methods: opportunities and challenges for novice researchers (during and beyond COVID-19), vol 23, pp 1–27
100. Good K, Shaw A (2021) Achieving a good impression: reputation management and performance goals. *Wiley Interdiscip Rev Cognit Sci* 12(4):e1552
101. Henrich J, Heine SJ, Norenzayan A (2010) Most people are not WEIRD. *Nature* 466(7302):29–29
102. Beran TN, Ramirez-Serrano A, Kuzyk R, Fior M, Nugent S (2011) Understanding how children understand robots: perceived animism in child–robot interaction. *Int J Hum Comput Stud* 69(7–8):539–550

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

Oshrat Ayalon is a human–computer interaction scientist with an interest in usable privacy and security, and adolescent online safety. Oshrat has a PhD in Industrial Engineering from Tel Aviv University. She did a postdoctoral fellowship at the Max Planck Institute for Software Systems. Oshrat is a Senior Lecturer (equivalent to Assistant Professor) at the Department of Information Systems at the University of Haifa.

Hannah Hok Hannah has a PhD in Psychology from the University of Chicago. She wrote her dissertation on how children think about the fairness of impartial decision-making, focusing on the flexibility of rules and their application. She is currently a visiting fellow at the Innovation Labs in the Harvard Library.

Alex Shaw. Alex has a PhD in Psychology from Yale University. He did a postdoctoral fellowship at the Booth School of Business. Alex is an Associate Professor at the University of Chicago and the director of the DIBS lab, which is focused on understanding how children and adults track other people's reputations and modify their behavior to create favorable impressions with others.

Goren Gordon PhD, PhD MBA. Goren has six academic degrees, a BA, MSc, and PhD in Quantum Physics, a BMSc, MBA, and another PhD in Neurobiology. He did his postdoc in MIT Media Lab's Personal Robots Group. Goren is the head of the Curiosity Lab in Tel-Aviv University and studies mathematical models of curiosity, implementing them in curious social robots and using them to assess and promote curiosity and other 21st-century skills in children. Goren is also the co-founder and Chief Science Officer of Curiosity Robotics, Ltd.