Works in Progress • Algorithmic Rights and Protections for Children

Authenticity and Co-design: On Responsibly Creating Relational Robots for Children

Marion Boulicault[1], Milo Phillips-Brown[2], Jacqueline M. Kory-Westlund[3], Stephanie Nguyen[4], cynthia breazeal[5]

[1]MIT, University of Adelaide, [2]Associate Professor, University of Oxford, [3]Ph.D.,MIT Media Lab; independent scholar with the Ronin Institute, writer, and artist, [4]Designer, Researcher, [5]Professor, MIT Media Lab

Published on: Jun 29, 2021

ABSTRACT

We at the MIT Personal Robots Group at the MIT Media Lab design and create *relational robots*—robots that form relationships with humans—to help children learn. In designing relational robots for children, we are, in effect, *designing relationships* between children and robots. To do so responsibly, we must address the pressing ethical question, "what kinds of relationships should we design?" Drawing from philosophy, disability studies, and our own empirical work, this paper considers this question by analyzing a frequently-voiced concern: that child–robot relationships are inevitably *in*authentic; perhaps they shouldn't be created at all. To address this concern and the more general question of what kinds of child–robot relationships we should design, responsible design requires working together with teachers, parents, children, and other stakeholders—it requires *co-design*. We describe how our group has collaborated with stakeholders from Boston-area schools to co-design relational robots.

Key Findings

- Designing relational robots for children is also *designing relationships* between children and robots. To do so responsibly, we must address the pressing ethical question, "what kinds of relationships should we design?"
- This paper considers the question of "what kinds of relationships should we design?" by analyzing the concern that child–robot relationships are inevitably *in*authentic.
- Responsible design requires working together with teachers, parents, children, and other stakeholders to *co-design* these sorts of relationships.

1. Introduction

Figure 1: A child with Tega

Meet Tega. Blue, fluffy, and AI-enabled, Tega is a *relational robot*: a robot designed to form relationships with humans. Created as a tool to improve early childhood education, Tega talks with children, plays educational games with them, engages in puzzle-solving, and participates in creative activities like making up stories and drawing. Tega uses AI algorithms to adapt to individual children's social, emotional, and curricular needs, thereby building a relationship with each child that keeps them engaged and improves how they learn. For example, Tega uses an algorithm that uses assessments of a child's language abilities to match the child with books to read that are at just the right language difficulty level to help build their vocabulary (Park et al., 2019).

For the past eight years, we at the Personal Robots Group at the MIT Media Lab have been developing and studying robots like Tega. Our hope is that relational robots can one day play a role in addressing urgent social issues, such as ensuring access to high quality early childhood education. To date, we have worked with over 400 preschool

and kindergarten children (ages 4 to 6 years) in diverse public schools. Our results so far are promising: we've found that children readily learn new words from robots like Tega (Kory-Westlund et al., 2017a), emulate the robot's phrases and vocabulary during storytelling activities (Kory-Westlund et al., 2017b; Kory-Westlund & Breazeal, 2019a), show more curiosity in response to a more curious robot (Gordon et al., 2015), and show more creativity when the robot models creative behavior (Ali et al., 2019). The closer the relationships between child and robot, the more effectively the child learns (Kory-Westlund & Breazeal, 2019b). In one of our studies, 49 children played one-on-one language-learning games with Tega once a week for eight weeks. The children who reported a closer relationship with Tega showed higher scores on language learning metrics, such as tests of vocabulary and ability to recall stories they'd heard or read (Kory-Westlund & Breazeal, 2019c; Kory-Westlund et al., 2018; Kory-Westlund, 2019). If designed and deployed responsibly, we believe that these robots have the potential to support teachers, communities, and governments in reaching education goals.

In designing relational robots for children, we are, in a sense, *designing relationships* between children and robots. To design the relational robots responsibly, then, we must address a pressing ethical question: "what kinds of child–robot relationships should we design?" That is, what should the relationship between a child and a robot like Tega be like? Or, what makes a good child–robot relationship?

This question has prompted discussion among our research participants and academics, and in the media (e.g. Turkle, 2007; Sparrow & Sparrow, 2006). Some of this discussion has centered on the concept of *authenticity:* good relationships are authentic relationships. A prominent concern is that child–robot relationships are inevitably inauthentic. That is, there's something inevitably inauthentic about *any* relationship that a child forms with a robot. If this is right, perhaps there's no way to responsibly design relational robots for children; perhaps we shouldn't be designing them at all (or, if we do, there must be *significant* benefit to outweigh the problem of inauthenticity).

In this paper, we focus on this concern of authenticity as a way of exploring the question of how to responsibly design relational robots like Tega.[1] We begin in section 2 by explaining what we mean by a 'relational robot', and expand on our motivations for building relational robots for use in early childhood education. In section 3, we analyze two different concerns about authenticity. Our analysis draws on our group's empirical research as well as on insights from philosophy and disability studies. In

section 4, we suggest a way forward. We argue that in order to design relational robots responsibly, it is ethically imperative that designers employ what's known as *co-design,* a framework that enlists stakeholders such as parents, teachers, and children themselves in answering the question: "what kinds of child–robot relationships should we design?" Using examples from our own research, we illustrate the significance of co-design for creating relational robots for children.

2. What are relational robots? And why build them?

2.1 What are relational robots?

Relational robots, we've said, are robots designed to form relationships with humans. They belong to a broader class of *relational technologies*, technologies that are designed to build relationships with humans. (This use of the term 'relational technology' dates back at least as far as Bickmore and Picard (2005).)

The idea that humans have relationships with technologies like robots is based on an understanding of a relationship—endorsed by various scholars[2]—on which humans can form relationships both with humans and with non-humans (with pets, for example).

This understanding of a relationship can be brought out by considering a related concept, that of a *social interaction*. A social interaction is commonly understood as an interaction between two agents whose behaviors are interdependent; the actions of one agent are responsive to the actions of the other (Berscheid & Reis, 1998). Social interactions include behaviors such as conversing, meeting another's gaze, taking turns, displaying emotion, gesturing, and performing what's known as behavior mirroring—matching one's behavior to that of the other. (The behaviors that make up social interactions are known as *social behaviors*.)

Many modern technologies have social interactions with humans—for example, entertainment robots like Aibo; personal home robots like Buddy, Jibo, and Mabu; and digital assistants like Alexa and Siri.[3]

Tega, too, socially interacts with humans—indeed, Tega is programmed to engage in a wide range of social behaviors. For example, Tega converses (using automatic speech recognition and by playing back recorded speech); meets the gaze of humans (e.g., Tega will "look" at the child's face when the child looks at it); and mirrors behavior (e.g. Tega will match the cadence of a child's speech or mirror a child's facial expressions). In our research, we've found that children tend to respond in kind. They readily converse with robots like Tega; mirror their behavior (e.g. mimic a robot's

facial expressions); take turns; share information about themselves; and help robots during joint activities (e.g., they turn the pages of a digital storybook for the robot and help the robot "practice" storytelling by re-telling stories). (Kanda et al., 2007; Kory-Westlund, 2019; Kory-Westlund & Breazeal, 2019a; Kory-Westlund et al., 2018; Park & Howard, 2015; Park et al., 2019; Serholt & Barendregt, 2016; Singh, 2018.)

Figure 2: A child with Tega

It takes more than just having a social interaction to be in a relationship, according to the understanding we're adopting. For example, if you meet the gaze of someone you pass on the street you do not thereby have a relationship with them; or if you ask Alexa what the weather will be tomorrow and you get a response, you do not thereby have a relationship with Alexa. Rather, relationships require a series of repeated, personalized, social interactions that can elicit feelings of *responsiveness* and *commitment* (as they're known in the literature).

Relationships unfold over time: in a relationship, repeated social interactions inform future social interactions. (Think of how your social interactions with a long-time friend differ from those with a stranger; this difference is partly due to a store of shared experiences.) In a relationship, you can refer back to activities shared in the past. Or, when you respond to someone, or some*thing*, with whom you have a relationship, you can in a sense *personalize* your response to them based on what you know from past interactions. As we noted in the introduction, this is precisely what Tega does. That is, Tega uses AI technology to tailor its future social interactions with a child on the basis of past interactions by, for example, picking books to read with children based on what it's learned about the child's literacy skills.

Feelings of responsiveness and commitment are umbrella terms that include positive feelings such as rapport, closeness, and attachment. Robots, of course, do not have feelings of responsiveness and commitment towards children; they don't have feelings at all! But children do. We've found in our studies that, for example, children report feeling as close to the robots as they feel to pets and favorite toys (Kory-Westlund et al., 2018). They readily say the robots are their friends (e.g, Kory, 2014) and frequently smile, laugh, and display various positive facial expressions when learning and playing with the robots (e.g., Kory-Westlund & Breazeal, 2019a; Kory-Westlund, 2019b).

Perhaps you are still skeptical that the word 'relationship' aptly describes the connections between children and relational robots. We explore skepticism of this kind in section 3.2. Ultimately, though, it's not essential to our purposes that child–robot relationships deserve the name. What is important is that children interact with robots in certain ways, and conceive of them in certain ways, that are similar in some respects to how they interact with and conceive of humans. It is the ethical dimensions of these interactions—not the label 'relationship'—that is our concern.

2.2 Why build relational robots for early childhood education?

Improving the quality and equity of early childhood education for all children is an issue of U.S. national educational importance (Hart & Risley, 1995). Early childhood is a critical time. It's when learning is most malleable and investments are most cost-effective for spurring long-term benefits to cognitive, academic, behavioral, and socioemotional outcomes (Heckman et al., 2010). A child who cannot read adequately in the first grade has a 90% probability of reading poorly in the fourth grade and a 75% probability of reading poorly in high school (Torgesen, 2004). Tragically, about one third of American children do not have access to high enough quality early

childhood education programs to prepare them to meet standards for kindergarten entry (Torgesen, 2004).

We at the MIT Personal Robots Group have designed technologies like Tega to help address some of these pressing social and educational issues facing our youngest learners. As we mentioned in the introduction, Tega is designed to help young children develop language and literacy skills and improve key learning attitudes such as curiosity, creativity, and the development of a growth mindset (the idea that one can develop one's talents and abilities through perseverance and effort (Dweck, 2008)). The use of AI technology to facilitate relationship-building between Tega and individual children makes Tega effective at meeting these goals when compared to non-relational technologies.[4] As such, Tega is well-positioned to support teachers in the classroom. For example, Boston-area preschool and kindergarten teachers from both private and public schools tell us that they would be excited to use robots like Tega during what they call "choice time"—a special time each day when children pick from a menu of different learning activities (Kory-Westlund et al., 2016).

Tega may also be effective at supporting parents and guardians with at-home learning (something that has become particularly urgent during the current COVID-19 pandemic). For example, research shows that children benefit from responding to dialogic questions (i.e. open-ended questions without clear right or wrong answers) (Hargrave & Sénéchal, 2000; Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1988). Tega is programmed to ask dialogic questions as a parent reads a book to a child, supporting the parent in facilitating learning (Boteanu et al., 2016; Chang et al., 2012; Nuñez, 2015).

Of course, issues concerning under-funding, support for teachers, and equitable access to high quality early childhood education are complex social issues, for which there can be no purely technical solution. Nevertheless, based on our research, we believe that technology like Tega has the potential to help address some dimensions of these issues—perhaps even in transformative ways.

3. Concerns about Authenticity

To design relational robots for children is to design relationships between robots and children. And so responsibly designing relational robots requires us to address the question, “what kinds of relationships should we design?”

As noted in the introduction, one widely-held answer to this question is based on the notion of authenticity: good relationships are (among other things) *authentic* relationships, and it's thus important that we design technologies for children that facilitate the creation of authentic relationships. During our studies, parents and teachers frequently raised the concern, in one form or another, that it's not possible for children to form authentic relationships with robots. This concern is echoed in the academic literature on relational robots: Sociologist Sherry Turkle, for example, insists that, in contrast to authentic human–human relationships, human–robot relationships are "superficial," "pretend," and "inauthentic" (Turkle, 2007). Philosophers Robert Sparrow and Linda Sparrow (2006) contrast human–robot relationships with "genuine" human–human relationships.[5]

Here in section 3, we analyze these concerns about authenticity. Our analysis reveals that there is no *one* unique authenticity concern; different ethical concerns go under the banner of "authenticity." We focus on two: the first is that child–robot relationships are *not real relationships* (section 3.2); the second is that these relationships are *deceptive* (section 3.3).

A note on the scope of our ambitions. First, we aren't aiming to analyze all possible concerns about authenticity. There are others that don't relate to either reality or deception. For example, Turkle (2007) argues that another reason that human–robot relationships are ethically alarming is that they may, in time, lead children to form inauthentic *human–human* relationships. Second, we are not advancing an analysis of what authenticity *is*. Rather, we aim to analyze two often-voiced concerns about child–robot relationships—concerns that have been stated in terms of authenticity—to better understand how to responsibly design relational robots.

3.1 On theorizing about authentic connections

Before investigating concerns about the authenticity of child–robot relationships, we'd like to step back and comment on theorizing about the authenticity of connections between humans and non-humans more broadly.[6] It is strikingly easy to make unjustified and potentially harmful assumptions about the inauthenticity of such connections—a fact that comes into relief with an example from disability studies.

Theologian Julia Watts Belser (2016) highlights a common assumption about the connections between disabled persons and assistive technologies, like wheelchairs: they are thought of as burdensome reliance, detracting from quality of life. Watts Belser illustrates this by pointing to the widely-used phrase 'wheelchair-bound,' which

evokes the idea of a wheelchair as something that "binds, traps, and constrains the human within its medicalized embrace" (2016, p. 6). On this view, disabled persons would be better off if they didn't have to rely on assistive devices.

Watts Belser's own experience as a wheelchair-user challenges this conventional thought. Rather than burdensome reliance, she sees her connection with her wheelchair as one of "intimate engagement between wheel and flesh that is central to my own embodied experience" (p. 7). The blogger Wheelchair Dancer echoes Watts Belser in describing her own connection with assistive devices: "My crutches are part of my arms—when I use them to make a dance line—and extra spines when I use them to support me and when I shift all of my weight on to the conjunction of arm and crutch." Wheelchair Dancer argues that we should conceptualize "disabled anatomy not as a set of functioning and failed body parts, bits that have partially been replaced by technology, but as a body that is extended and expanded by its technology" (quoted from Watts Belser, p. 12). The connection between Wheelchair Dancer and her assistive technology is extensive, expansive, and empowering.

Once we consider Watts Belser's and Wheelchair Dancer's perspectives, it's hard to think of an adequate definition of authenticity that would label as inauthentic their connections with their wheelchairs and crutches. And yet this is the opposite of what we'd expect if we adopted the conventional—and, to many, seemingly obvious—understanding of how disabled persons relate to assistive technologies, an understanding that is based on problematic ableist assumptions.

Of course, the relationships between children and robots are both practically and ethically different in significant ways from the connections between disabled persons and assistive devices. Children don't, for example, usually think of robots as extensions of their bodies. And while child–robot relationships may face a certain stigma, that stigma cannot be compared to the ableist oppression that persons with disabilities face. Nonetheless, a lesson can be drawn from scholars working in disability studies: if we're theorizing about what counts as an authentic connection or relationship, we must be epistemically humble, which is to say that we cannot put too much weight behind our own thoughts and intuitions. We must look to those who have direct knowledge—or what's known as "lived experience"—of the connection or relationship. The judgments that may come easily must be carefully critiqued and interrogated. We ought to take extra caution with new types of relationships, like relationships between children and AI-enabled relational robots, where conventional wisdom may not apply.

3.2 Inauthenticity as unreality?

With that in mind, let us turn to the concerns raised about the authenticity of child–robot relationships. In our research, we've found that when some study participants—such as teachers and parents—express concerns about authenticity, they sometimes seem to be expressing a concern that the relationship a child forms with a robot is somehow *unreal*, or at least *less real*, than the relationship a child forms with a teacher or friend. One could reconstruct this concern as follows. Human–human relationships are real; indeed human–human relationships set the ideal for what a real relationship is. Any relationship that lacks the qualities of human–human relationships is a mere approximation of a real relationship. It is less than real, and therefore inauthentic.

This thought has intuitive appeal. Although human–robot relationships have various qualities found in paradigmatic human–human relationships (see section 2), they lack many others. Today's robots do not empathize with a child who has stubbed her toe; they do not feel joy if a child writes them a thoughtful note; they do not care if they never again see a child with whom they've interacted. One would be quick to label "inauthentic" human–human relationships that lack these qualities: imagine someone who claims to be your friend but who doesn't empathize with you, is not moved by a thoughtful note, or wouldn't care if they never saw you again; this is *not a real friend*.

But we suggest that it's hasty to leap to the conclusion that *any* kind of relationship —especially human–non-human relationships—is fake or unreal if it lacks certain qualities, such as the ability to empathize. Is your relationship with your dog, for example, not real if he is indifferent to a thoughtful note? Presumably not. Human–human relationships don't set the standard for *all* relationships. Rather, we submit, there are relationships of different kinds, each of which might have different standards of "realness" or authenticity. What makes your relationship with a friend authentic, for example, is not, intuitively, the same as what makes your relationship with your dog authentic.

If this idea is right, then human–robot relationships may constitute "real" relationships —just a different kind of real relationship than human–human relationships. We've observed evidence of this in our research. We found that children generally do not conceive of robots as equivalent to their human peers and caregivers, or even as the same as their pets, toys, or computers (Kory-Westlund 2019; Kory-Westlund & Breazeal, 2019a; Kory-Westlund et al., 2018).

This finding is well illustrated by a study we conducted to gauge how children perceive Tega. We asked children to complete a sorting activity in which we presented them with pictures of different entities including a frog, a cat, a baby, a robot from a movie (like R2D2 from the *Star Wars* films), a mechanical robot arm, Tega, and a computer (Kory-Westlund & Breazeal, 2019c). Children were asked to place these pictures on a spectrum with an human adult on one extreme, and a table on the other. Children frequently placed Tega near the middle of the spectrum, between a computer and a cat, indicating that they saw Tega as more human-like than a computer but less human-like than a cat (which they generally placed closer to the adult than Tega). In another study—which we referenced in section 2.1—we asked children to talk about how close they felt to Tega in comparison to pets, toys, friends, and parents. On average, children said that they felt similarly close to Tega as to their pets and favorite toys, but less close than how they feel to friends and parents (Kory-Westlund et al., 2018). These data lend credence to the thought that child–robot relationships needn't be, or needn't necessarily be, a less real, approximate version of human–human relationships. Child–robot relationships may simply be a different kind of relationship, with their own distinct standards of 'realness.'

In other words, we're suggesting that the fact that child–robot relationships lack qualities of human-human relationships does not mean—as some have worried—that child–robot relationships are less real and therefore inauthentic. There is evidence, for example, that children consider robots to be a different kind of entity than humans, suggesting that child–robot relationships may likewise be of a different kind than human–human relationships. Child–robot relationships may plausibly have their own distinct standards of realness and authenticity. As such, it does little to simply charge that child–robot relationships are 'unreal' without specifying a standard of 'realness' or 'authenticity' against which to judge the relationships.

Nonetheless, we don't think that the inauthenticity-as-unreality concern is entirely misguided. The issue is how it's been *expressed*. When theorists and our research participants say that they are concerned about unreality, we think they are most charitably understood as giving voice to a different concern. The concern is that child–robot relationships are somehow *off* or *not quite right*. In other words, child–robot relationships are—for a reason not so easily articulated by unreality—not the kinds of relationships we should be designing for our children. (It is not only unreal relationships that are problematic. Think, for example, of a child's relationship with a bully: this isn't a relationship that a child should be in, but that has nothing to do with unreality. It may be all too real!)

The inauthenticity-as-unreality concern seems to bring us right back where we started: “what kinds of child–robot relationships should we design?” (Or should we even be designing them at all?) Inauthenticity-as-unreality doesn’t help answer this driving question, since it doesn’t say what standards of ‘realness’ we should be judging the relationships against. In section 4, we will offer a way to address this driving question that we argue is more effective than considerations of realness. But before that, we first consider another commonly-raised concern about the authenticity of child–robot relationships, this one having to do with deception.

3.3 Inauthenticity as deception?

According to a second authenticity concern, child–robot relationships are inauthentic not because they are unreal, but because they are *deceptive*. Some relational robots are programmed to represent themselves—in some sense or other—as empathetic, curious, or as having any number of emotions or mental states. For example, Tega can mirror children's facial expressions, giving the appearance of an emotional reaction; or, when playing a learning game, Tega can say things like "Ooh!" while leaning forward and opening its eyes wide, giving the appearance of curiosity. Other robots we’ve designed, such as one named Green the DragonBot, explicitly ascribe themselves emotions, saying, for example, "I like playing with you!"

The concern is that in behaving in these ways, robots—or, more accurately, the robot’s designers and programmers—may lead children to believe (wrongly) that the robots are capable of emotion (Picard & Klein, 2002; Sparrow & Sparrow, 2006; Turkle, 2007). This *inauthenticity-as-deception* concern can be understood in various ways (see Coeckelbergh (2012) for a taxonomy of these various ways). Here, we articulate one version of the concern.

The idea that deceptive relationships are inauthentic is familiar from everyday life. If you learned that your partner has lied to you for decades about their real name; pretended to love you when they did not; or only cared about your relationship insofar as it served their professional aims, all of this would be not only hurtful but would indicate something about the relationship itself. A relationship built on deception can rightly be called inauthentic (at least to some extent and in certain cases).

3.3.1 Are children wrong about what robots are like?

The concern that child–robot relationships are deceptive presupposes that children are indeed mistaken about what robots are like. But are they? Do children mistakenly

believe that today's relational robots—like Tega—have attributes, like a capacity for emotion, that they do not in fact have?

Children *do* ascribe emotions to relational robots. They say things about robots like "She's kind," "if you just left him here and nobody came to play with him, he might be sad," and "he likes sharing stuff, like stories" (Kory-Westlund et al., 2018). One child, when asked what he would do if one of our robots was sad, suggested he would "buy ice cream to make him happy, robot ice cream" (Kory, 2014). But of course these robots lack the capacity to feel kind or sad; they lack the capacity to like; if they were given ice cream—whether robot or human ice cream—it would not make them feel anything at all.

One conclusion to draw is that children are indeed mistaken about what robots are like. We'd like to counsel caution about accepting this conclusion too readily. First, as we noted in section 3.2, children tell us that they don't conceive of robots as equivalent to friends, parents, or other humans. This may suggest that while children use words like "sad" to describe robots, they may conceive of the sadness that they ascribe to robots differently than the sadness they'd ascribe to a friend or a parent. Just as a child conceives of a robot eating "robot ice cream" rather than "human ice cream," so too might she think of a robot as having "robot feelings" rather than "human feelings."

Second and most obviously, it's uncontroversial that children engage in make-believe games and play activities where they knowingly pretend that things are other than what they are. It's something that adults do with children—pretending, for example, that a Winnie the Pooh bear or Furby is alive and has feelings. All of this is considered an important and positive childhood activity. It's not a stretch to see Tega playing a similar role to these toys. Indeed, we've found in our research that parents and teachers pretend that Tega has feelings. Given that children aren't "deceived" by a Winnie the Pooh bear or Furby, we shouldn't be too quick in thinking they're deceived by Tega.

3.3.2 What do inauthenticity-as-deception concerns mean for the design of relational robots?

One could plausibly argue that Tega and toys like Winne the Pooh and Furby differ when it comes to deception. Tega does many things that such toys do not, like sustain conversations with children and match their facial expressions and the pace and cadence of their speech. And most distinctively, Tega collects data from children and uses AI technology to personalize and adapt its interactions over time. As this AI

technology advances, it's easy to imagine that Tega-like robots of the future will behave in ways that leave children genuinely believing that robots have thoughts and feelings.

If this is the case now or in the future—i.e. if child–robot relationships are or will be somehow deceptive—would that be a cause for concern? We'll argue that the answer to this question is not straightforward.

Adults frequently deceive children—or don't disabuse them when they're mistaken about certain things, like whether their pet has died, whether the tooth fairy exists, or whether their dinner contains vegetables. The ethical implications of such deception differs considerably from deception of adults. Compare: a parent sneaking vegetables into a child's dinner and telling them there are no vegetables, versus a company doing the same with their employees. We may imagine that in both cases, the deception leads to an outcome that benefits the deceived; with the child and parent, though, the deception has a different moral complexion than with the employee and company.

Using relational robots does not, as we see it, raise some *distinctive* or *new* concern over and above those about deception of the presence of vegetables in dinner or the existence of imaginary beings. Rather, it seems clear that in general, parents, teachers, and others who serve care-taking roles can use limited deception for the benefit of children—that is, deception in certain select cases and to certain select ends. And using relational robots promises to be of the exact kind that warrants such limited deception: helping the child to develop intellectually and emotionally. As we noted above, our research indicates that relational robots indeed help children learn.

More generally, deception seems to fall into a broad category of behavior whose moral status depends on whether the recipient is an adult or a child. While in many cases it would be wrong to *control* the life of an adult—e.g. deciding what she eats, who she can socialize with, or what her bed time is—such treatment is not only appropriate for children, but is the responsibility of care-takers. Deception is a certain way of controlling a person.

This is not to say that *all* control of children is good; and in particular, not to say that all deception of children is good. Far from it. Our point is rather that the moral import of deceiving children is complex. With children, we cannot simply equate "deceptive relationship" with "a relationship a child should not have" (nor can we equate it with "a relationship a child *should* have"). To evaluate the ethical import of deceiving a child, one needs to know more, as philosophers have argued. In particular, one needs

to know the *context* in which the deception is taking place. For instance, one needs to know: *why* is the child being deceived? (See e.g., Pallikkathayil (2019).) Is it to facilitate learning? to eat more vegetables? to spend more money on toys? And *who* is doing the deceiving? (See e.g., White (ms).) A parent? robot? teacher? corporation?

The overarching question in need of an answer is "what kinds of relationships should we design?" According to the most straightforward understanding of the inauthenticity-as-deception concern, any deceptive relationship is problematic; if child–robot relationships are deceptive, that is automatically cause for concern. But as this section (3.3.2) showed us, things are not so clear-cut. Some deceptive relationships may be problematic, while others may not be. Simply pointing to deception (just like simply pointing to the notion of unreality) is insufficient to tell us which relationships we should design. To tell whether deception in a child–robot relationship is problematic, we need to know the context—the *who, when,* and *why* of the deception. This is all to say that we need to know the context surrounding the child–robot relationship to answer the question: "what kinds of child–robot relationships should we design?"

4. Responsible design with authenticity in mind: an argument for co-design

We've said that in designing relational robots for children we are, in effect, designing relationships. This is because children will form different kinds of relationships with different kinds of robots. For example, whether a robot says that it feels certain ways, or how it responds to a child asking "Do you love me?" may affect whether the relationship is deceptive (and so, according to some, inauthentic).

In section 3, we argued that the two authenticity concerns we considered don't take us far enough in determining the kinds of child–robot relationships we should design, or whether we should be designing such relationships at all. In this section, we offer a more promising path forward. Rather than aiming to identify a fixed definition of the kind of child–robot relationship we should be designing (e.g. giving a definition of an authentic relationship), we focus on the *process* by which we answer the question, "what kinds of child-robot relationships should we design?" More specifically, we'll argue that this question can be answered responsibly only if it is answered collaboratively, using a family of methodologies known as collaborative design, or *co-design*.[7]

Section 4.1 explains the spirit and methods of co-design. Section 4.2 argues that co-design is imperative for addressing the question, "what kinds of child–robot

relationships should we design?" And section 4.3 shows co-design of child–robot relationships in action: we describe how we at the MIT Personal Robots Group have used co-design methods in designing our relational robots.

4.1 What is co-design?

Co-design, most simply, is design in *partnership* with the people and communities who are or might be affected by a given technology. (As is common, we'll call these people and communities *stakeholders*.) Co-design overlaps with related approaches known as "participatory design," "human-centered design," and "inclusive design;" and indeed, it is often used as an umbrella term for these approaches. Costanza-Chock (2020) offers a useful encapsulation of co-design as "the full inclusion of, and accountability to, and control by, people with direct lived experience of the conditions [that] designers […] are trying to change" (p. 26). *And Also Too*, a design studio dedicated to co-design, describes their work as "guided by two core beliefs: first, that those who are directly affected by the issues a project aims to address must be at the center of the design process, and second, that absolutely anyone can participate meaningfully in design." (And Also Too, n.d.).

What does it mean to design in partnership with stakeholders? To answer this question, it is helpful to contrast co-design with *user research* methods, which aim to obtain information from stakeholders. For example, a designer creating a meditation phone application might conduct focus groups with potential users to learn what these stakeholders want and how they might interact with such an application. User research methods provide information, but it is up to the designers what they do with that information. For example, the application designers might use what they learn to ensure that the app helps users meet their own meditation goals. Or they might use the information in order to design the application to maximize the time a user spends on it (regardless of the users' own goals and values).

Co-design is different. While user research methods might form an important *part* of a co-design approach, these methods alone are not sufficient for co-design. This is because co-design requires that stakeholders be included not only as *sources* of information, but also as *decision-makers*. If we were using co-design to design a meditation app, stakeholders would not only provide information to the designers; they would be partners in making design decisions.

There is no one-size-fits-all approach to co-design; rather, co-designers use a variety of methods and strategies for including stakeholders as design partners, depending on

the nature of the project and on the specific stakeholders. These might include participatory technology assessments (Banta, 2009; Hennen, 2012), citizen juries (Gooberman-Hill et al., 2008; Street et al., 2014), and global interdisciplinary observatories (Hurlbut, 2018). (For more details on these methods, see Sample et al. (2019).) There are also co-design methods specifically targeted towards children. Druin (2002), for example, articulates a framework where children can take a variety of roles in the broader design process of new technologies—that of user, tester, informant or design partner. This framework emphasizes that all partners "must acknowledge that a child has the right to partake and possess an active role" in the design process.

Co-design is not new to the design of relational robots. A research team at UC San Diego used co-design methods in the design of robots for dementia caregiving. They conducted a six-month long community design–research process, built relationships with members of local community centers, and empowered caregivers by collaborating with them in the design of physical prototypes (Moharana et al., 2019). Other research teams have adopted co-design methods in designing relational robots for children. For example, researchers have explored the use of Cooperative Inquiry methods with intergenerational teams in the design of social robots for children (Arnold et al., 2016). This approach allows groups of children across age ranges, with different levels of knowledge and learning styles, to explore new information together. Researchers in the Netherlands and the United Kingdom working on designing robots for children with autism implemented co-creation sessions with children, family members, and professionals affiliated with autism spectrum disorder (Huijnen et al., 2017). To facilitate collaboration and trust among participants, the sessions were held in environments familiar to participants, who sat in a "U-shape" arrangement (as opposed to in rows, for example) so that they could look at each other while speaking.

The need for facilitating trust brings up one of the central challenges—and promises—of co-design. We live in a world with extreme social inequities and hierarchical power structures, illustrated forcefully by the growing power divides between the technology sector and the rest of society. It may be difficult to find ways to effectively include stakeholders as partners, especially those who have been historically excluded from design processes, such as those from low-resourced or otherwise marginalized communities. For instance, in the context of relational robots for children, family members from low-resourced communities may not have access to transportation or have the time or resources to attend co-creation sessions or lab meetings. In addition, stakeholders from marginalized groups may not trust the universities or corporations

building these technologies. This is why a co-design approach requires accounting for stakeholder histories and power dynamics.

4.2 The case for co-design in building relational robots for children

Why is co-design necessary for designing relational robots responsibly? As we just discussed, co-design says that to responsibly design any given technology, the design process must include those people and communities who are affected by the technology. The primary motivation behind co-design is a matter of *justice:* those affected by a technology deserve a say in how they will be affected. In other words, stakeholders of any given technology deserve a say in how that technology is designed (see e.g. Costanza Chock (2020)). We'll argue for something more specific: that stakeholders of relational robots deserve a say in answering the question, "what kinds of child–robot relationships should we design?"

Outside of the context of relational robots, the question of what kinds of relationships children should have is the province of parents, teachers, children themselves, caregivers, communities, etc.—or rather it is their province within certain bounds, on which more shortly. It is not the province, or not the sole province, of traditional designers of technologies. Why would things be any different with the question of which relationships children should have with relational robots? As co-design dictates, a broad range of stakeholders—not just product designers and researchers—need power over decisions about the kinds of child–robot relationships that children have.

To make the point more concrete, think about one of the authenticity concerns we examined in section 3—specifically, that child–robot relationships are deceptive (and thereby inauthentic). We argued that simply knowing that a child–robot relationship is deceptive (if it's deceptive at all) isn't enough to determine whether it's a relationship that children should or should not have. Deception may be problematic in certain contexts, but not in others. One determining factor in whether deception is problematic is *who decides* to deceive the child. Imagine that a food corporation creates a snack for children without disclosing to the public that it contains vegetables. Imagine further that you have a child who buys this snack and eats it. She has been deceived, and, it seems, in a problematic way. The problem is not that it's never okay to deceive children about the contents of their food. It could be fine for *you* (the parent) to trick your child into eating vegetables. Rather, the problem—or at least part of the problem—is that it is not the place of a corporation to decide on its own whether to deceive children. As a parent, you deserve to have a say in whether your child is deceived. A similar thing goes, we maintain, for if and when child–robot

relationships should be deceptive. It is not the place of traditional designers to decide this matter alone; parents and other stakeholders deserve a say, too.

We don't mean to suggest that if parents, teachers, or other stakeholders think that it's appropriate to deceive a child, then they are thereby correct. There are, as we've said, simply cases where children should not be deceived. (For example, if parents deceive their child without regard to her interests.) More generally, there are certain kinds of relationships—e.g. abusive or oppressive relationships—that children should *never* have, regardless of whether parents, teachers, a community, or anyone else thinks that they should. This sets a certain boundary on what child–robot relationships we should be designing. But within this boundary, the question remains: "what kinds of child–robot relationships should we design?" This question, we've argued, is for co-design to answer.

4.3 An example of co-designing relational robots with diverse stakeholders

In section 4.1, we outlined the concept of co-design, and in section 4.2 we argued that co-design methods are imperative for the responsible design of relational robots for children. Here in section 4.3, we offer an example, based on our work designing Tega and Green the Dragonbot, of what it looks like in practice to apply co-design methods to the design of relational robots for children.

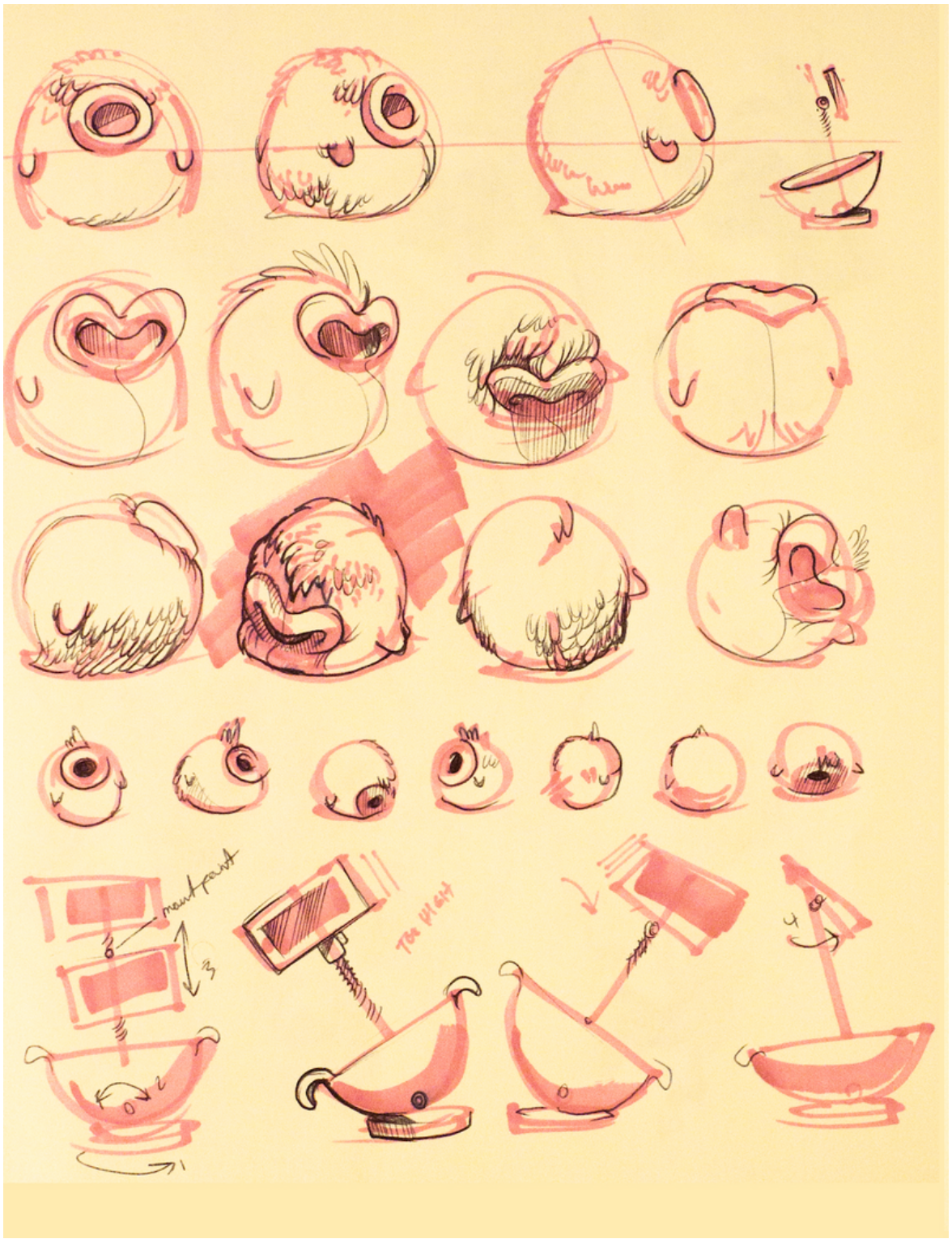

Figure 3: Concept sketches from the early design phase of Tega

First, some background on our stakeholders and our co-design methods. The stakeholders with whom we engaged included parents, teachers, school administrators, early childhood development experts, and children from Boston-area public schools that serve households from a variety of socioeconomic backgrounds. We made special efforts to include stakeholders from ethnically and linguistically diverse backgrounds, including bilingual and English-language-learning children and families. We used a variety of methods, including meetings, surveys, interviews, and focus groups to learn about stakeholders' values and perspectives on the use of relational robots in early childhood education. Our co-design methods were iterative: we'd have discussions with stakeholders, go back to our lab to integrate their perspectives and values into our design work, come back to the stakeholders for more discussion and feedback, and so on.

We also developed co-design methods specifically aimed at children. We brought children and parents *together* into the lab to interact with lower-fidelity relational robot prototypes (i.e., prototypes that did not include all the features we might deploy in a robot in a school). These were often remotely controlled by a person (as opposed to being autonomous)—this method is known as *Wizard of Oz*.[8] This prototyping method helped us understand the types of emotional interactions children would want to have with a robot, and, crucially, it helped us do so *before* we built any AI algorithms that powered child–robot interactions completely autonomously. We also developed simple games and picture-based questionnaires for children, like the sorting activity discussed in section 3.2. In one questionnaire, we asked children about their perceptions of Tega's social and relational attributes (e.g., "Let's pretend Tega didn't have any friends. Would Tega not mind or would Tega feel sad?", "Does Tega really like you, or is Tega just pretending?"); children could point to pictures of Tega in their responses as well as explain their thinking. We invited children to draw pictures to different prompts, including many concerning potential relationships they might form with Tega, such as "draw a picture of your dream robot and what you do together."

Co-design approaches had material impacts on how we built robots like Tega. In our early discussions with stakeholders, we identified a widely held assumption: parents and teachers frequently assumed that robots like Tega would take the role of a teacher, i.e. that Tega would relate to children as a source of authority and expertise. (This is not surprising given that many research labs and companies are developing intelligent expert tutoring systems, like Squirrel AI, AutoTutor, and COLit.[9]) However, when we talked with parents, teachers, and children about how they *wanted* Tega to relate to children, we heard a different message. Many believed that children's educational

needs would be better served if relational robots were to take the role of a *peer-like learning companion* as opposed to an expert teacher (e.g. Chen et al. (2020)).

Stakeholders offered a variety of reasons for preferring a peer-like robot over a teacher-like robot. Teachers explained that that they saw value in a robot that could be used as a “motivator or reinforcer,” provide a “non-judgmental safe learning space,” and introduce children to “activities they might not otherwise do” (Kory-Westlund et al., 2016) —all things they believed would be more easily achieved with a robot in a peer-like role. Teachers also expressed concerns that if the robot were to take on a teacher-like role, children would perceive it as competing with human teachers in the classroom. Further, teachers worried that a teacher-like robot might be more likely to “replace” teachers in the future; this, teachers believed, would harm how children learned, and could result in teachers losing their jobs. Children, too, expressed a preference for engaging in peer-like relationships with robots. They responded more positively to a robot that asked them to play as another child would ("Do you want to play a story game?") than a robot that directed the activity in a teacher-like way ("Let's practice our storytelling now"). Children also reacted positively and learned more effectively when robots appeared friendly and inviting, like a special kind of pet, rather than a distant authority figure. Children favored plush fabrics and bright contrasting colors, often petting the robot, or putting their arm around it as they played games together.

Figure 4: A child with Green the DragonBot

In light of this, we adjusted our designs: rather than designing the robot as an expert teacher, we cultivated a child–robot relationship by designing Tega to be a peer-like or pet-like learning companion. For example, we programmed Tega to use language that is more child-like (and less teacher-like), such as the language mentioned in the previous paragraph. We also designed Tega to occasionally make mistakes—e.g. Tega sometimes incorrectly answers questions about vocabulary or the content of a story—to make it appear less authoritative (and also to allow it to model a growth mindset; see page 9). Based on children's interactions and preferences, we chose bright, soft material and a cute, animal-like design so that the robot would look like a kind of special, friendly pet.

These design choices had the intended effect. We observed that children in our studies tended to relate to the robot as a pet or playmate (Kory-Westlund, 2019; Kory-Westlund & Breazeal, 2019a, 2019b; Kory-Westlund et al., 2018; Park et al., 2019). They

assumed the robot liked playing with them, too: "I know Green [the robot] likes to play with me, so I know he's happy!" (Kory, 2014).

When we invited parents and guardians to participate in co-design sessions, we made further discoveries about what kinds of child–robot relationships we should design. We learned that many parents wanted to be involved as their children learned with Tega. We thus designed Tega to engender a *group relationship* among children, robots, and adults. For example, we created a special French language-learning activity for Tega, and asked 16 families to participate in the activity to hear their feedback and perspectives. As part of the activity, the robots only used French words when conversing with children. Parents participated in the learning activity by pointing out (in English) to the child when the robot was using new words and then prompting the child to repeat or use that word: "How do you say ‘bye’ in French?" (Freed, 2012) The robot facilitated French learning by indirectly prompting the parent to engage in guiding and teaching their child. Parents told us that they experienced a *socially inclusive* experience, contrasting it with what they saw as socially exclusive experiences they have when their child uses a tablet (like an iPad). It would not have been possible to understand the importance and value of these group relationships without the close collaboration with parents and guardians as co-designers.

5. Conclusion

In an interview in *The Guardian*, Sherry Turkle warns that “if people start to buy the idea that machines are great companions […], as they increasingly seem to do, we are really playing with fire” (quoted in (Adams, 2015)). We agree with Turkle that developing relational robots raises genuine social and ethical concerns. But we also believe that, when designed and implemented responsibly, these technologies have the potential to yield significant benefits and transformative change. We’ve argued that to responsibly build relationships between children and robots, and to address concerns about authenticity, co-design is required. Stakeholders deserve a say in deciding what kinds of child–robot relationships we should design. If we want to “avoid playing with fire,” all of us need to be in this together.

Author Bios

Marion Boulicault (MIT, University of Adelaide) is a graduate student in the Philosophy Department at MIT. Starting in 2021, she will be a Lecturer at the University of Adelaide. Her research focuses on feminist philosophy of science and technology.

Milo Phillips-Brown is an Associate Professor of Philosophy at the University of Oxford Faculty of Philosophy and Department of Computer Science, a Tutorial Fellow at Jesus College, and a Senior Research Fellow in Digital Ethics and Governance at the Jain Family Institute. His research is about the ethics of technology, ethical engineering pedagogy, and the philosophy of mind and language.

Jacqueline M. Kory-Westlund is an independent scholar with the Ronin Institute, writer, and artist with a PhD from the MIT Media Lab. Her research has focused on using social robots to support and engage young children in learning, including how children understand social robots, how children's relationships connect to their learning, especially over long-term interactions, and the ethics of using robots in children's lives.

Stephanie Nguyen is a human-computer interaction designer and researcher, specializing in data privacy, user experience design, and tech policies that impact vulnerable populations. She is an appointed member of IEEE Standards Association's Global Advisory Council on Children's Experiences and Trustee at 5Rights Foundation focused on children's digital rights and data literacy.

Cynthia Breazeal is a professor of Media Arts and Sciences at the Media Lab. Her research is about social robots and their responsible use, spanning technical innovations, user-centered design, and the psychology of engagement with particular focus on education, health, aging and wellness application domains.

References

Adams, T. (2015, October 18). Sherry Turkle: 'I am not anti-technology, I am pro-conversation.' *The Guardian*. http://www.theguardian.com/science/2015/oct/18/sherry-turkle-not-anti-technology-pro-conversation

Ali, S., Moroso, T., & Breazeal, C. (2019). Can children learn creativity from a social robot? *Proceedings of the 2019 on Creativity and Cognition*, 359-368.

And Also Too (n.d.). *What is co-design?* Retrieved December 8, 2020 from www.andalsotoo.net/what-is-codesign

Arnold, L., Lee, K. J., & Yip, J. C. (2016). Co-designing with children: An approach to social robot design. *Proceedings of the 11th AMC/IEEE International Conference on Human–robot Interaction*.

Banta, D. (2009). What is technology assessment? *International Journal of Technology Assessment in Health Care, 25*(S1), 7-9.

Berscheid, E., & Reis, H. T. (1998). Attraction and close relationships. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology, 1-2*(4), (pp. 193–281). McGraw-Hill.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293-327.

Boteanu, A., Chernova, S., Nunez, D., & Breazeal, C. (2016). Fostering parent–child dialog through automated discussion suggestions. *User Modeling and User-Adapted Interaction*, *26*(5), 393–423.

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science*, 1–11.

Chang, A., Breazeal, C., Faridi, F., Roberts, T., Davenport, G., Lieberman, H., & Montfort, N. (2012). Textual tinkerability: Encouraging storytelling behaviors to foster emergent literacy. *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, 505–520.

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*, *150*, 103836.

Coeckelbergh, M. (2012). Are emotional robots deceptive? *IEEE Transactions on Affective Computing*, *3*(4), 388–393.

Cole, R., Wise, B., & Van Vuuren, S. (2007). How Marni teaches children to read. *Educational technology, 47*(1), Special issue: Pedagogical agents, 14-18.

Costanza-Chock, S. (2020). *Design justice.* MIT Press.

Csikszentmihalyi, M., & Halton, E. (1981). *The meaning of things: Domestic symbols and the self*. Cambridge University Press.

Design Justice Network, (n.d.). *Design justice network principles*. Retrieved December 8, 2020 from https://designjustice.org/read-the-principles

Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018). How smart are the smart toys?: Children and parents' agent interaction and intelligence attribution. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–240.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, 21, 1-25.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Ballantine Books.

European Group on Ethics in Science and New Technologies (2018). *Statement on artificial intelligence, robotics and 'autonomous' systems.* Retrieved December 8, 2020, from https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382

Fink, J., Mubin, O., Kaplan, F., & Dillenbourg, P. (2012). Anthropomorphic language in online forums about Roomba, AIBO and the iPad. *Advanced Robotics and Its Social Impacts (ARSO), 2012 IEEE Workshop On*, 54–59.

Freed, N. A. (2012). *"This is the fluffy robot that only speaks French": Language use between preschoolers, their families, and a social robot while sharing virtual toys* [Master's Thesis]. Massachusetts Institute of Technology.

Friedman, B., & Hendry, D. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press.

Friedman, B., Kahn, P. H., Jr., & Hagman, J. (2003). Hardware companions?: What online AIBO discussion forums reveal about the human–robotic relationship. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 273–280.

Gooberman-Hill, R., Horwood, J. and Calnan, M. (2008), Citizens' juries in planning research priorities: process, engagement and outcome. *Health Expectations*, 11, 272-281. doi:10.1111/j.1369-7625.2008.00502.x

Gordon, G., Breazeal, C., & Engel, S. (2015). Can children catch curiosity from a social robot? *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human–Robot Interaction*, 91–98.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies, 14*(3), 575-599.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, *15*(1), 75–90. https://doi.org/10.1016/S0885-2006(99)00038-1

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children*. Paul H Brookes Publishing.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. & Yavitz, A. (2010). The rate of return to the high/scope perry preschool program. *Journal of Public Economics,* 94(1-2), 114–128.

Hennen, L (2012). Why do we still need participatory technology assessment?. *Poiesis & Praxis* 9, 27–41. https://doi.org/10.1007/s10202-012-0122-5

Huijnen, C. A. G. J., Lexis, M. A. S., Jansens, R., & de Witt, L. P. (2017). How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents and professionals. *Journal of Autism and Developmental Disorders, 47*, 3079–3096.

Hurlbut, J. B., Jasanoff, S., Saha, K., Ahmed, A., Appiah, A., Bartholet, E., Baylis, F., Bennett, G., Church, G., Cohen, I. G., Daley, G., Finneran, K., Hurlbut, W., Jaenisch, R., Lwoff, L., Kimes, J. P., Mills, P., Moses, J., Park, B. S., Parens, E., … Woopen, C. (2018). Building capacity for a global genome editing observatory: Conceptual challenges. *Trends in Biotechnology*, *36*(7), 639–641.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*, first edition. Retrieved December 8, 2020, from https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

Kahn, P. H., Friedman, B., & Hagman, J. (2002). I care about him as a pal: Conceptions of robotic pets in online aibo discussion forums. *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, 632–633.

Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human robot interaction. *IEEE Transactions on Robotics, 23*(5), 962–971.

Kelley, H. H., Berscheid, E., Christensen, A., Harvey, J., Huston, T., Levinger, G., McClintock, E., Peplau, L., & Peterson, D. (1983). *Close relationships*. Freeman.

Kory, J. (2014). *Storytelling with robots: Effects of robot language level on children's language learning* [Master's Thesis]. Massachusetts Institute of Technology.

Kory-Westlund, J. M. (2019). *Relational AI: Creating long-term interpersonal interaction, rapport, and relationships with social robots.* [PhD Thesis]. Massachusetts Institute of Technology.

Kory-Westlund, J. M., & Breazeal, C. (2019a). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI, 6*(81). doi:10.3389/frobt.2019.00081

Kory-Westlund, J. M., & Breazeal, C. (2019b). Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Frontiers in Robotics and AI, 6*(54). doi:10.3389/frobt.2019.00054

Kory-Westlund, J. M., & Breazeal, C. (2019c). Assessing children's perceptions and acceptance of a social robot. *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 38–50.

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

Kory-Westlund, J. M., Gordon, G., Spaulding, S., Lee, J. J., Plummer, L., Martinez, M., Das, M., & Breazeal, C. (2016). Lessons from teachers on performing HRI studies with young children in schools. *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 383–390.

Kory-Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., DeSteno, D., & Breazeal, C. (2017b). Flat versus expressive storytelling: Young children's learning and retention of a social robot's narrative. *Frontiers in Human Neuroscience*, *11*.

Kory-Westlund, J. M., Park, H. W., Williams, R., & Breazeal, C. (2018). Measuring young children's long-term relationships with social robots. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 207–218.

Moharana, S., Panduro, A. E., Lee, H. R., & Riek, L. D. (2019). Robots for joy, robots for sorrow: Community based robot design for dementia caregivers. *2019 14th ACM/IEEE International Conference on Human–Robot Interaction (HRI)*, 458-467.

Montreal Declaration Responsible AI (n.d.). *The declaration.* Retrieved December 8, 2020, from https://www.montrealdeclaration-responsibleai.com/the-declaration

Nuñez, D. S. (2015). *GlobalLit: A platform for collecting, analyzing, and reacting to children's usage data on tablet computers* [Thesis, Massachusetts Institute of Technology]. Massachusetts Institute of Technology.

Pallikkathayil, Japa (2019). The truth about deception. *Philosophy and Phenomenological Research*, *98*(1), 147-166.

Park, H. W., & Howard, A. M. (2015). Retrieving experience: Interactive instance-based learning methods for building robot companions. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 6140–6145.

Park, H. W., Gelsomini, M., Lee, J. J., & Breazeal, C. (2017). Telling stories to robots: The effect of backchanneling on a child's storytelling. *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, 100–108.

Park, H. W., Grover, I., Spaulding, S., Gomez, L., & Breazeal, C. (2019, July). A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 687-694.

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., & Breazeal, C. (2017). Growing growth mindset with a social robot peer. *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, 137–145.

Picard, R. W., & Klein, J. (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting With Computers*, *14*(2), 141–169.

Riek, L. D. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human–Robot Interaction*, *1*(1).

Sample, M., Boulicault, M., Allen, C., Bashir, R., Hyun, I., Levis, M., Lowenthal, C., Mertz, D., Montserrat, N., Palmer, M. J., Saha, K., & Zartman, J. (2019). Multi-cellular

engineered living systems: Building a community around responsible research on emergence. *Biofabrication*, *11*(4).

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, what's up?": A mixed-methods studies of in-home conversational agent usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868.

Serholt, S., & Barendregt, W. (2016). Robots tutoring children: longitudinal evaluation of social engagement in child–robot interaction. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction, 64*, 1-10. https://doi.org/10.1145/2971485.2971536

Singh, N. (2018). *Talking Machines: Democratizing the design of voice-based agents for the home.* [Master's Thesis]. Massachusetts Institute of Technology.

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, *16*, 141-161.

Street, J., Duszynski, K., Krawcxyk, S. & Braunack-Mayer, A. (2014). The use of citizens' juries in health policy decision-making: A systematic review. *Social Science & Medicine*, *109*, 1-9. https://doi.org/10.1016/j.socscimed.2014.03.005.

Torgesen, J. K. (2004). Preventing early reading failure—and its devastating downward spiral. *American Educator*, *28*(3), 6–19.

Turkle, S. (2005). *The second self: Computers and the human spirit*. MIT Press.

Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction Studies*, *8*(3), 501–517.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Hachette UK.

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day care. *Developmental Psychology, 28*(6), 1106–1114. https://doi.org/10.1037/0012-1649.28.6.1106

Watts Belser, J. (2016). Vital wheels: Disability, relationality, and the queer animacy of vibrant things. *Hypatia, 31*(1), 5–21.

Weiss, A., Wurhofer, D., & Tscheligi, M. (2009). “I love this dog”—Children’s emotional attachment to the robotic dog AIBO. *International Journal of Social Robotics*, *1*(3), 243–248.

White, P. Quinn. *Honesty and discretion.* University of Nebraska, Lincoln.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, *24*(4), 552–559. https://doi.org/10.1037/0012-1649.24.4.552

Wise, B., Cole, R., Van Vuuren, S., Schwartz, S., Snyder, L., Ngampatipatpong, N., Tuantranont, J., & Pellom, B. (2005). Learning to read with a virtual tutor: Foundations to literacy. *Interactive Literacy Education: Facilitating Literacy Environments Through Technology*, 31–75.

Footnotes

1. Authenticity is not the only ethical issue related to the responsible design of relational robots. Others include concerns about data collection, social injustices around access to technology, privacy and security, corporate power, and the future of work, to name just a few. In this paper we focus on authenticity concerns. ↩
2. See, for example, Berscheid & Reis, (1998); Csikszentmihalyi & Halton, (1981); Kelley, et al., (1983). ↩
3. For more on Buddy see http://www.bluefrogrobotics.com/; on Jibo, https://www.jibo.com/; on Mabu, http://www.cataliahealth.com/; on Alexa, see Sciuto et al. (2018). For academic work on Aibo, see e.g. Fink, et al., (2012); Friedman, et al., (2003); Kahn, et al., (2002); Weiss, et al., (2009). ↩
4. For a representative sample of work see Breazeal et al., (2016); Chen et al., (2020); Kory-Westlund, (2019); Kory-Westlund & Breazeal, (2019a, 2019b); Westlund et al., (2017); Park, et al., (2017). ↩
5. Other researchers and scholars have also weighed in on the question of authenticity, e.g. Coeckelbergh (2012); Picard & Klein (2002). See also additional work by Turkle (2005, 2017). ↩
6. We’re using ‘connection’ as a general term that encompasses relationships. ↩

7. We don’t mean to suggest that co-design is the only appropriate or useful methodology for the responsible design of relational robots. The responsible design of any technology requires many complementary approaches, including those related to legal compliance, monitoring and assessment, and data governance. For details of other approaches, see, for example: the Montreal Declaration for Responsible AI (n.d.); the IEEE’s recommendations on ethically-aligned designed design (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019); the European Group on Ethics in Science and New Technologies statement on artificial intelligence, robotics and 'autonomous' systems (European Group on Ethics in Science and New Technologies, 2018); and value-sensitive design (e.g. Friedman & Hendry (2019)). ↩

8. Wizard-of-Oz is a common technique enabling researchers to explore aspects of interaction not yet backed by autonomous systems. See (Riek, 2012). ↩

9. For more details on these systems, see squirrelai.com/, ace.autotutor.org/IISAutotutor/index.html, (Cole, et al., 2007), and (Wise, et al., 2005). ↩