Deception, Secrets, Children, and Robots: What's Acceptable?

Jacqueline Kory Westlund
MIT Media Lab
20 Ames St.
Cambridge, MA 02139
jakory@media.mit.edu

Cynthia Breazeal
MIT Media Lab
20 Ames St.
Cambridge, MA 02139
cynthiab@media.mit.edu

ABSTRACT
This short paper describes two scenarios that occurred during human-robot interaction studies with children involving (1) telling secrets to robots, and (2) deception about the robot's agency. These stories raise questions pertaining to children's privacy, the importance of trust in a relationship, the deception inherent in Wizard-of-Oz studies, and children's general construal of robots. Many of these questions remain relevant as we move toward autonomy, and when we consider other vulnerable populations.

Categories and Subject Descriptors
I.2.9 [Artificial Intelligence]: Robotics---Commercial robots and applications; J.4 [Computer Applications]: Social & Behavioral Sciences---Psychology; K.4.1 Ethics.

General Terms
Design, Human Factors, Theory.

Keywords
Children; deception; ethics; long-term interaction; privacy; social assistive robotics; social robots.

1. INTRODUCTION
Using a Wizard-of-Oz setup in human-robot interaction (HRI) studies presents several ethical issues regarding deception and privacy. At the most basic level, the human interacting with the remote-operated robot is deceived into thinking the robot is acting autonomously. They may disclose sensitive information to the robot that they would not tell a human, not realizing that a human is hearing everything they say. They may feel betrayed when they find out about the deception. Given that social robots are designed to draw us in, often engaging us emotionally and building relationships with us [2], the robot itself could be deceptive in that it appears to have an emotional response to you but "in reality" does not [3]. These issues remain as we move toward autonomy.

Here, we present two stories about child-robot interactions that highlight some questions about privacy, trust in human-robot relationships, and deception about robot agency that may arise when robots take the roles of tools, teachers, and companions in children's lives. We discuss children in particular because they readily treat robots as friends and companions [6–8], and are often the target population in HRI studies.

HRI'15 Workshops, Mar 2, 2015, Portland, OR, USA.

2. TELLING SECRETS

2.1 A Story
In an interview at the end of a study, we asked children if they would tell a secret to the robot. The children had played a storytelling game designed to support language learning with the robot eight times over the past two months (it was a Wizard-of-Oz setup, primarily to deal with speech recognition; see [8]). As indicated by other interview questions, they generally thought of the robot as a social being and a friend. Many said yes: they would tell the robot a secret. One child said, "Yeah, she couldn't tell anyone else because she's a robot." Related work has shown similar results [4], including that children are just as willing to tell a secret to a robot as to an adult [1].

2.2 Questions
It was a Wizard-of-Oz study. Obviously, the human controlling the robot would hear what the child said to it, secret or no. If the child did disclose a secret, how should it be treated? Should the human pretend to not know to maintain the illusion that the robot is a separate entity? If the secret pertains to bullying, abuse, or another situation involving potential danger to a person, the experimenter is obligated by IRB protocols to report it. But how should less severe secrets be treated? This seems the bigger issue, particularly for longer interactions, in which the child's trust in the robot may be important to the interaction's longer-term goal, such as in education [5]. A robot's "betrayal" of the child's trust could negatively affect this relationship, and thus, negatively impact the desired outcome (of, e.g., learning with the robot). One solution is for experimenters to take care not to reveal their "extra" knowledge except in the extreme cases involving potential harm to persons. This may not be far removed from the usual precautions taken during a study to maintain the Wizard-of-Oz illusion. But, as will be discussed in the next sections, the deception inherent in Wizard-of-Oz studies is its own issue.

Autonomous robots present the same problems. A human may review the data collected by the robot during an interaction with a child, and learn a secret, or even just view potentially sensitive information. In the context of a study, we assume the experimenters will view all the data. However, if we move from lab studies to field studies in schools and homes, what extra precautions should we take during data collection for data protection? What rights to privacy should a child have?

3. DECEPTION ABOUT AGENCY

3.1 A Story
A couple years ago, we took one of our robots to a preschool for a show-and-tell day. Each child was invited to speak to the robot and share information about their favorite animal; the robot

(which was remote-operated by a human) replied with an interesting fact about that animal. Later, we showed the children what the robot looked like on the inside and talked about how robots are made. We showed them that we had been remote-controlling this robot, and invited each child to trigger one of the robot's facial expressions using our remote-operation interface. After all this, one child insisted that he be allowed to teach the robot how to make a paper airplane. He later announced to the human operator that he had taught the robot about airplanes. Somehow, a disconnect existed between what he had just learned about the robot and the robot's human operator, and the character or social agent that he perceived the robot to be.

3.2 Questions

This incident threw light on an intriguing question: How are children actually construing robots? This question has been addressed to an extent in papers about children's beliefs about robots as agents (e.g., [6, 9]), though the questions asked tend to assume that the robot is autonomous or at least that if a Wizard-of-Oz scenario used, the children are unaware of it. The additional question raised here is this: If we tell children that we are controlling the robot, what does that mean to them? Can they understand that the robot and the human behind it are, in some sense, the same? That, e.g., if they tell the robot a secret, the human will know, too? The child mentioned in the story above was told that the robot was controlled by a person, but still acted as if the robot was its own separate entity. Does a child's understanding depend on their developmental age, or, perhaps, their understanding of theory of mind? We could probe children's conceptions of robots through a developmental 2x2 study in which children ages 2-8 interact with either an autonomous or remote-operated robot, in which they are told either that the robot is acting on its own or that a person is puppeting the robot. Analyzing children's behavior, such as how they speak to the robot or about the robot, their trust in it as an informant or a friend, and their theory of mind capabilities may give us insight into how these children construe robots.

Wizard-of-Oz setups raise their own questions. When, if ever, is it okay to deceive children about the nature of robots? How, at the end of a study, after a child asks if the robot will be joining the rest of the preschool at the preschool's end-of-year picnic, do you gently explain that the robot is not, in fact, their friend whom they've played with for the past eight weeks – it's just a machine, controlled by a human? How do you balance the desire to maintain children's imagination or fantasy about a robot, versus following the moral directive not to deceive? After all, children are frequently deceived about other agents to preserve some mystery or fascination – how many children do you know who believe in Santa Claus? How many think that the Cinderella encountered at Disneyland *really is* Cinderella, not just a person in costume? (A worthwhile question to ask is whether *these* deceptions are acceptable ones, given their pervasiveness in our culture.) If the illusion is upheld, how might the child's experience with this robot color their expectations about the capabilities of other robots and other technology? Or, if our aim is for the robot to be, say, a tutor, and if the educational outcome is greater if the child believes the robot is its own agent, is it still acceptable to deceive the child about the robot's autonomy, even past the end of a study?

Additional questions about a child's relationship with the robot arise at the end of a study. Children may think of the robot as a companion and friend. What happens when the robot goes back to the lab? The children generally do not see the robot again. How might children perceive this end of the relationship – especially if they had spent many weeks playing with the robot? What kinds of stories should we provide to children about why the robot left – e.g., is it appropriate to tell children that the robot is moving away or attending a different school (as they might have had other friends who left for similar reasons)? Little research has investigated this issue. We could gain some initial insight through interviews conducted after longer-term studies have concluded – a week later, as well a year or more later. If we plan to introduce social robots into people's lives, we need to understand what kind of impact they will have. Will they impact us in ways similar to a human friend? Or will they be more like a beloved pet, a favorite childhood toy, or something new?

4. CONCLUSION

Current studies of HRI raise important ethical questions about how we should design and curate robots in children's lives, both for single encounters and longer interactions. Many questions highlighted through Wizard-of-Oz setups about deception, privacy, and trust remain relevant for autonomous systems, while the questions raised by considering child-robot interactions can be extended to include any vulnerable population. This paper does not attempt to answers all these questions. However, acknowledging the unique ethical concerns that robots bring to light is a first step in figuring out how to tackle them.

5. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) under Graduate Research Fellowship Grant No. 1122374. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not represent the views of the NSF.

6. REFERENCES

[1] Bethel, C.L., Stevenson, M.R. and Scassellati, B. 2011. Secret-sharing: Interactions between a child, robot, and adult. *Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2489–2494.

[2] Breazeal, C.L. 2004. *Designing Sociable Robots*. MIT Press.

[3] Coeckelbergh, M. 2012. Are Emotional Robots Deceptive? *IEEE Transactions on Affective Computing*. 3, 4, 388–393.

[4] Fior, M., Ramirez-Serrano, A., Beran, T., Nugent, S. and Kuzyk, R. 2010. Children's relationships with robots: Robot is child's new friend. *Journal of Physical Agents*. 4, 3, 9–17.

[5] Harris, P.L. 2007. Trust. *Developmental Science*. 10, 135–138.

[6] Kahn, P.H., Kanda, T., Ishiguro, H., Freier, N.G., Severson, R.L., Gill, B.T., Ruckert, J.H. and Shen, S. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology*. 48, 2, 303.

[7] Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction.* 19, 1, 61–84.

[8] Kory Westlund, J. and Breazeal, C. 2015. The interplay of robot language level with children's language learning during storytelling. *Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction - Extended Abstracts* (Portland, OR).

[9] Severson, R.L. and Carlson, S.M. 2010. Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks*. 23, 8, 1099–1103.