Children's Overtrust: Intentional Use of Robot Errors to Decrease Trust

Denise Y. Geiskkovitch, James E. Young

***Abstract*—Robots are being developed to help in educational settings (among others) with young children. Research suggests that children may overtrust robots, which can have a negative impact. We suggest the use of intentional, egregious robot errors as one technique to mitigate such overtrust. Additionally, how robots attempt to recover from intentional and unintentional errors could also help in reducing children's trust towards them. In this paper we provide our reasoning behind the proposed purposeful use of errors, as well as suggestions for how various types of errors could be used to decrease trust towards robots.**

I. Introduction

Young children (i.e., below 10 years of age) are now commonly exposed to robots, be it in public places such as shopping malls or airports, or increasingly in their daily lives in schools and at home. Research suggests that children perceive robots differently than adults do; while adults are aware that social robots are machines (even if they treat them socially), children are more likely to see robots as being alive, and interact with them as social beings [1]. Children even tend to attribute moral rights to robots [2], and ascribe them with affective abilities [3] and free will [4].

Research suggests that children are trusting of robots (e.g., [5], [6]), and can exhibit technological or interpersonal trust towards them [6]. Young children (~3 years old) are in general very trusting of both robots and people, more so than other age groups, and by age 7 children tend to trust robots more than they do adults in some situations [5]. However, such trust may not always be warranted or desired, as several concerns exist over privacy, security, and inappropriate learning.

We believe that one of the ways in which researchers and designers might be able to mitigate the effects of overtrust in children is through the use of designed, on-purpose robot errors. Although robots are extremely error-prone, the typical approach is for researchers and designers to focus on disguising or recovering from errors and malfunctions, to enhance people's perceptions and interactions with the robot. We argue for the opposite, to intentionally leverage robot errors in order to lower children's trust towards robots.

II. How Robot Errors Affect Children's Trust

A robot's errors can influence children's perceptions and trust of the robot. Accuracy errors, for example, tend to decrease young children's trust towards agents and robots [7], [8], while providing at least some correct responses can increase trust [9]. Responsiveness errors (when it takes a robot a long time to respond) can also lower children's trust in robots [10]. This research points to the potential for robot errors to be used intentionally to decrease young children's trust in robots.

Figure 1. Child observes as a robot makes an error, considering whether to trust it or not. (image used with parental consent)

III. Using Errors to Mitigate Trust

Robots can produce a number of errors that may successfully, and intentionally, reduce children's trust towards them [11]. Below we include some examples of how researchers could employ these errors to prevent children's overtrust of robots.

A. Robot Responsiveness

Robot responsiveness, defined as the amount of time that it takes for the robot to produce an appropriate response, be it social, verbal, or physical, can affect children's trust [10]. For example, a robot that takes 20 seconds to respond, as opposed to 5, may be perceived as faulty or inexperienced, altering trust towards it. Research suggests that children trust robots that show response delays less than those that do not, pointing to the potential for researchers to utilize responsiveness to alter trust.

1) Proposed techniques

Utilizing intentional response delays might help researchers and designers to lower young children's trust towards robots. Response delays could be implemented in a random fashion, causing an occasional delay in the robot's speech or actions to encourage children to consider the robot's true (limited) abilities, liveness, and intelligence. Delays may also be included or exaggerated in specific high-impact contexts, to aim to dilute perceptions of robot perfection. For example, a robot tutor may be able to quickly provide answers to math questions, but purposefully have delays when responding to social or off-topic questions.

When utilizing response delays, care must be taken to maintain desired trust (i.e., in the task that the robot is intended

D. Y. Geiskkovitch and J. E. Young are with the University of Manitoba, Winnipeg, MB R3T2N2 Canada (corresponding author is Denise Y. Geiskkovitch: denise.geiskkovitch@cs.umanitoba.ca).

to complete) while decreasing undesired trust (beyond the robot's task or abilities).

B. Robot Accuracy

How accurate or reliable a robot is in the information it provides and tasks it completes can affect how much young children trust it [8]. A robot that provides clearly incorrect information, such as labelling a ball as a chair, is unlikely to be trusted. Similarly, children's trust is likely to be low in a robot that appears to be incompetent at completing a task. While accuracy in our sense encompasses several aspects of reliability, correctness, and precision, these are all errors that could take place in the presence of children, influencing their perceptions. Accuracy, or lack thereof, could therefore be utilized to decrease children's trust towards robots.

1) Proposed techniques

While classic engineering approaches tend to strive towards robots being as accurate as possible, we believe that with the proper techniques, intentionally programming a robot to be strategically inaccurate could benefit child-robot interaction. For example, a robot could be programmed to occasionally provide incorrect responses to questions, or failed attempts at walking or grasping objects if that is not its main purpose. However, careful consideration is necessary to ensure the reduction of undesired trust, while still maintaining desired trust. This technique might be especially helpful in decreasing trust while maintaining rapport; research suggests that while accuracy errors lead to lowered trust in adults, the robot was also perceived as more likeable [12]. While similar studies have yet to be conducted with young children, this approach may enable researchers and designers to lower trust while potentially maintaining positive perceptions, therefore leading to better outcomes with children.

C. Appropriate Recovery Strategies

When a robot does make an error, whether intentionally designed or not, how the robot continues that interaction (i.e., its trust recovery strategies) can be further used to impact trust. Techniques such as providing explanations of why the error occurred and apologizing have been found to assist trust rebuilding [13]. These strategies could be used to ensure that desired trust towards a robot is maintained when necessary. However, this approach could also be utilized to decrease undesired trust, by highlighting the occurrence of errors and the robot's true abilities. This strategy might aid in decreasing children's trust towards robots, which does not always occur when there are errors [8], if the errors are highlighted or explained.

1) Proposed techniques

When robots provide explanations of why errors occurred, statements about their lack of abilities or reasons for the error could be included to foster distrust or regain trust, depending on the design goal. For example, a robot could say "Sorry, I made a mistake. I don't always know everything" to highlight its true abilities, and therefore decrease children's trust towards it in undesired cases. On the other hand, a robot could instead say "Sorry, I made a mistake. I couldn't see the question properly" which suggests that the robot is capable of performing the task properly, and therefore regain trust in it. This approach could encourage children to be mindful of when they should place their trust on robots, and how much trust is appropriate. This transparency in a robot's abilities can sometimes decrease children's trust towards it [14].

IV. Conclusion

Children tend to perceive robots as a mix between humans and machines, possessing the best of both worlds, and leading to high levels of trust towards them. Although trust may be sought after in some situations, overtrust in robots could lead to unwanted or unsafe situations for children. In this paper we posit the idea of utilizing robot errors to decrease children's trust towards robots. While this is a new, and somewhat controversial, area of research, we believe that this approach could be beneficial to children's relationships with robots, as well as their overall social development.

References

[1] A. S. Rao and M. P. Georgeff, "BDI agents: From theory to practice," in *Proceedings of the 1st International Conference on Multiagent Systems*, 1995, pp. 312–319.

[2] P. H. Kahn *et al.*, "'Robovie, you'll have to go into the closet now': Children's social and moral relationships with a humanoid robot," *Dev. Psychol.*, vol. 48, no. 2, pp. 303–314, 2012.

[3] T. N. Beran, A. Ramirez-Serrano, R. Kuzyk, M. Fior, and S. Nugent, "Understanding how children understand robots: Perceived animism in childrobot interaction," *Int. J. Hum. Comput. Stud.*, vol. 69, no. 7–8, pp. 539–550, 2011.

[4] T. N. Beran, A. Ramirez-Serrano, O. G. Vanderkooi, and S. Kuhn, "Reducing children's pain and distress towards flu vaccinations: A novel and effective application of humanoid robotics," *Vaccine*, vol. 31, no. 25, pp. 2772–2777, 2013.

[5] C. Di Dio *et al.*, "Shall I trust you? From child–robot interaction to trusting relationships," *Front. Psychol.*, vol. 11, no. April, pp. 1–14, 2020.

[6] C. L. Van Straten, J. Peter, R. Kühne, C. De Jong, and A. Barco, "Technological and interpersonal trust in child-robot interaction: An exploratory study," in *Proceedings of the 6th International Conference on Human-Agent Interaction*, 2018, pp. 253–259.

[7] J. H. Danovitch and R. Alzahabi, "Children show selective trust in technological informants," *J. Cogn. Dev.*, vol. 14, no. 3, pp. 499–513, 2013.

[8] D. Y. Geiskkovitch, R. Thiessen, J. E. Young, and M. R. Glenwright, "What? That's not a chair!: How robot informational errors affect children's trust towards robots," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, 2019, pp. 48–56.

[9] Z. Henkel *et al.*, "He can read your mind: Perceptions of a character-guessing robot," in *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*, 2017, pp. 242–247.

[10] C. Breazeal, P. L. Harris, D. Desteno, J. M. Kory Westlund, L. Dickens, and S. Jeong, "Young children treat robots as informants," *Top. Cogn. Sci.*, vol. 8, no. 2, pp. 481–491, 2016.

[11] D. Y. Geiskkovitch and J. E. Young, "Social robots don't do that: Exploring robot-typical errors in child-robot interaction," *ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 200–202, 2020.

[12] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction," in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 2016, pp. 493–500.

[13] S. Tolmeijer *et al.*, "Taxonomy of trust-relevant failures and mitigation strategies," *Proceeding 15th ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 3–12, 2020.

[14] C. L. van Straten, J. Peter, R. Kühne, and A. Barco, "Transparency about a robot's lack of humanlike psychological capacities: Effects on child-robot perception and relationship," *ACM Trans. Human-Robot Interact.*, vol. 9, no. 2, pp. 1–22, 2020.