Clarifying Social Robot Expectation Discrepancy

Developing a Framework for Understanding How Users Form Expectations of Social Robots

James M. Berzuk
Department of Computer Science
University of Manitoba
Winnipeg, Canada
berzukj@myumanitoba.ca

James E. Young
Department of Computer Science
University of Manitoba
Winnipeg, Canada
young@cs.umanitoba.ca

ABSTRACT

When engaging with a social robot, people form expectations about the robot that may not align with its real behaviour and abilities. This gap is known as expectation discrepancy, and can confuse and disappoint users. We are developing a framework that can be used to understand and compare instances of expectation discrepancy between robots by considering the sources of those expectations. In doing so, we aim to provide a structure and unified vocabulary that can be used to support description and comparison of robot designs and the expectations users form of them. We have begun by examining theoretical work on expectations in interactions between people, and are working to synthesize this into an initial foundation. We will then refine this into a final social robot expectation framework by conducting a survey of expectation formation and discrepancy in existing social robots and projects.

CCS CONCEPTS

• Human-centered computing → HCI theory, concepts and models

KEYWORDS

human-robot interaction; social robots; expectation discrepancy; framework; survey

ACM Reference format:

James M. Berzuk and James E. Young. 2023. Clarifying Social Robot Expectation Discrepancy: Developing a Framework for Understanding How Users Form Expectations of Social Robots. In *Proceedings of the 2023m ACM/IEEE International Conference on Human-Robot Interaction (HRI'23), March 13-16, 2023, Stockholm, Sweden. ACM, New York, NY, USA, 3 pages.* https://doi.org/10.1145/3568294.3580078

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

HRI'23, March 13–16, 2023, Stockholm, Sweden.

© Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9970-8/23/03...$15.00.
DOI: https://doi.org/10.1145/3568294.3580078

1 INTRODUCTION

Social robots are designed to simplify interaction with people through human or animal-like designs that people can readily understand [4]. However, in interacting with social robot designs, people can form a plethora of expectations of the robot, based on that design: for example, a person may reasonably assume that a robot with hands and fingers can pick up items. Of course, the robot may not have this capability, creating an *expectation discrepancy* [21] where people may not only misunderstand how to interact with a robot, but may be disappointed by it, impacting the success of the interaction [10].

The expectations people form of social robots can emerge from a range of sources, including decades of fanciful depictions in media [5,20], and are heavily influenced by the robot's own design [12,14]. Outwardly human-like design features, while effective in promoting familiarity [4] and empathy [18], may lead to expectations of human-like capability. While it may not be feasible to meet such high expectations, some have instead suggested exploring ways to design robots that more accurately imply their capabilities [11,15]. By clearly setting expectations in advance, the robot may be able to avoid disappointing the user when they see the robot's actual abilities.

We are working to support engagement with this issue through the development of a framework that allows for understanding and comparison of instances of the problem. While various works have investigated expectation formation and discrepancies, the field will benefit from more holistic and systematic examination of the subject.

We are developing this framework by first establishing a theoretical grounding in work on expectation formation in human-human interactions. We are synthesizing these works from a social robotics perspective, and will use the results to survey expectations-related features in robots in the field, ultimately producing a descriptive framework that can be used to summarize and compare instances of expectation formation and discrepancy.

Beyond supporting greater understanding and discussion, this framework will serve as a foundation for the development of applicable design tools in the future, informing designers on what features they may add and remove from their robot in order to achieve more accurate expectations in users.

2 RELATED WORK

Numerous works have explored different aspects of expectations surrounding robots. One focus of such works is to examine the sources of initial expectations, be they depictions in media [5,20] or features of the robot itself [14,17]. Another major focus is to examine how users respond to gaps between their expectations and actual robot capabilities [10], and more generally how they respond to robots' errors and failures [13,16,19]. Attempting to address the issue, some works have investigated ways to moderate user expectations to better suit the robot's real abilities [11,15]. While significant efforts have been made in this direction, there is much to gain with more structure and a unified vocabulary, supporting a more cohesive approach toward to the problem.

As a young field, human-robot interaction is still exploring many areas, and an important step for any issue is to map out the problem space. By synthesizing observed trends across the field into a systematized framework, we can obtain a better understanding of the issue and its key components. This approach has been highly successful in mapping out other areas of the field, including more general, high-level frameworks designed to describe whole instances of human-robot interaction [2,22], identifying more specific components and patterns that make up an interaction [7,9], or considering particular aspects of interaction such as dialogue [3]. We are now applying this approach to support the understanding and analysis of robot expectation formation and discrepancy.

3 METHODOLOGY

We are currently undergoing the first steps in the development of this new framework. Our approach begins with collecting and reviewing existing literature on expectation formation in a human-human context, which we will synthesize into an initial set of framework dimensions. We will then iteratively refine these dimensions through a survey of robots in the field, and will evaluate the work using case studies selected from this survey.

3.1 Theoretical Foundations

We draw from expectation formation in human-human work to ground our exploration of expectations of social robots. For example, there are several theories regarding expectations in interactions between humans which may be relevant to robots as well, such as expectancy violations theory [6] and simulation theory [8]. This will also include considering particular psychological phenomena, such as the *pratfall effect* (people perceived as highly capable become more likeable after making a mistake, while people who are already perceived as unreliable become less likeable after making the same mistake) [1] which has received prior consideration in a robotic context [13].

We will synthesize this work from the perspective of the unique demands and needs of social robots in comparison to people (e.g., limitations of computation, online connectivity). From this investigation, we will identify patterns and areas of common interest, and use these to unify the ways in which people are discussing the issue. We will also collect a list of features known to impact expectations, and put these together to assemble an initial working framework.

3.2 Synthesis and Analysis of Existing Robots

We will refine the framework to suit the needs of actual robots by conducting a survey of the field. We will survey past works featuring robots that were published at HRI, looking for details about the outward features of the robot and any information about expectations and responses that participants had.

Using these collected systems, we will perform an open-coding thematic analysis, classifying each robot according to our framework and iteratively updating the dimensions as we proceed. This will result in a framework that reflects both theoretical human-human literature and present robots in the field.

3.3 Case Studies

We will select out of our survey three to four current key robots and use them as case studies. We will use the framework to summarize and compare them, in order to demonstrate the frameworks' descriptive ability.

3.4 Contributions

There are two key contributions that will result from our work. The first is the descriptive framework we will have developed in order to support greater understanding of robot expectations. The second is a quantitative summary resulting from our survey, in which we will be able to show the frequencies of expectation-relevant design features in the field. This will provide an overview of common approaches and designs in contemporary systems.

4 FUTURE WORK

The framework we are developing will not only enable greater understanding of the issue, but we also plan to use it as a foundation for future work.

The next step following the development of the framework will be to quantitatively examine the impact of the various features of the framework by conducting a participant study. We will present participants with different robot designs, adjusting features along the different dimensions to evaluate how they affect what participants expect from the robot.

Beyond this, we are interested in building on those results to develop an applicable design tool that can inform designers about how they can adjust their designs in order to achieve the expectations they desire for their robots.

ACKNOWLEDGMENTS

This project was funded by the Natural Sciences and Engineering Research Council of Canada, through their Discovery Grants program. This project was further supported by the University of Manitoba Graduate Fellowship.

REFERENCES

[1] Elliot Aronson, Ben Willerman, and Joanne Floyd. 1966. The effect of a pratfall on increasing interpersonal attractiveness. *Psychon Sci* 4, 6 (June 1966), 227–228. DOI:https://doi.org/10.3758/BF03342263

[2] Christoph Bartneck and Jodi Forlizzi. 2004. A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog*

No.04TH8759), IEEE, 591–594. DOI:https://doi.org/10.1109/roman.2004.1374827

[3] James M Berzuk and James E Young. 2022. More Than Words: A Framework for Describing Human-Robot Dialog Designs. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '22), IEEE Press, 393–401.

[4] Cynthia Breazeal. 2003. Toward sociable robots. *Robotics and Autonomous Systems* 42, 3–4 (March 2003), 167–175. DOI:https://doi.org/10.1016/S0921-8890(02)00373-1

[5] Ulrike Bruckenberger, Astrid Weiss, Nicole Mirnig, Ewald Strasser, Susanne Stadler, and Manfred Tscheligi. 2013. The Good, The Bad, The Weird: Audience Evaluation of a "Real" Robot in Relation to Science Fiction and Mass Media. In *Social Robotics* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 301–310. DOI:https://doi.org/10.1007/978-3-319-02675-6_30

[6] Judee K. Burgoon and Jerold L. Hale. 1988. Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs* 55, 1 (March 1988), 58–79. DOI:https://doi.org/10.1080/03637758809376158

[7] Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2016. Human-robot interaction design using interaction composer: Eight years of lessons learned. In *ACM/IEEE International Conference on Human-Robot Interaction* (HRI '16), IEEE Press, 303–310. DOI:https://doi.org/10.1109/HRI.2016.7451766

[8] Robert M. Gordon. 1992. The simulation theory: Objections and misconceptions. *Mind & Language* 7, 1–2 (1992), 11–34. DOI:https://doi.org/10.1111/j.1468-0017.1992.tb00195.x

[9] Peter H. Kahn, Nathan G. Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina H. Ruckert, Rachel L. Severson, and Shaun K. Kane. 2008. Design patterns for sociality in human-robot interaction. In *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots* (HRI '08), Association for Computing Machinery, New York, NY, USA, 97–104. DOI:https://doi.org/10.1145/1349822.1349836

[10] Takanori Komatsu, Rie Kurosawa, and Seiji Yamada. 2012. How Does the Difference Between Users' Expectations and Perceptions About a Robotic Agent Affect Their Behavior? *Int J of Soc Robotics* 4, 2 (April 2012), 109–116. DOI:https://doi.org/10.1007/s12369-011-0122-y

[11] Minae Kwon, Sandy H. Huang, and Anca D. Dragan. 2018. Expressing Robot Incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, Chicago IL USA, 87–95. DOI:https://doi.org/10.1145/3171221.3171276

[12] Minae Kwon, Malte F. Jung, and Ross A. Knepper. 2016. Human expectations of social robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. DOI:https://doi.org/10.1109/HRI.2016.7451807

[13] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI* 4, (2017). Retrieved August 31, 2022 from https://www.frontiersin.org/articles/10.3389/frobt.2017.00021

[14] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20), Association for Computing Machinery, New York, NY, USA, 33–42. DOI:https://doi.org/10.1145/3319502.3374839

[15] Steffi Paepcke and Leila Takayama. 2010. Judging a bot by its cover: an experiment on expectation setting for personal robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (HRI '10), IEEE Press, Osaka, Japan, 45–52.

[16] Marco Ragni, Andrey Rudenko, Barbara Kuhnert, and Kai O. Arras. 2016. Errare humanum est: Erroneous robots in human-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 501–506. DOI:https://doi.org/10.1109/ROMAN.2016.7745164

[17] Samantha Reig, Jodi Forlizzi, and Aaron Steinfeld. 2019. Leveraging Robot Embodiment to Facilitate Trust and Smoothness. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 742–744. DOI:https://doi.org/10.1109/HRI.2019.8673226

[18] Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (HRI '09), Association for Computing Machinery, New York, NY, USA, 245–246. DOI:https://doi.org/10.1145/1514095.1514158

[19] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *Int J of Soc Robotics* 5, 3 (August 2013), 313–323. DOI:https://doi.org/10.1007/s12369-013-0196-9

[20] Eduardo Benitez Sandoval, Omar Mubin, and Mohammad Obaid. 2014. Human Robot Interaction and Fiction: A Contradiction. In *Social Robotics* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 54–63. DOI:https://doi.org/10.1007/978-3-319-11973-1_6

[21] Lena T. Schramm, Derek Dufault, and James E. Young. 2020. Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design. *ACM/IEEE International Conference on Human-Robot Interaction* Figure 2 (2020), 439–441. DOI:https://doi.org/10.1145/3371382.3378280

[22] H.A. Yanco and Jill Drury. 2004. Classifying human-robot interaction: an updated taxonomy. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, IEEE, 2841–2846. DOI:https://doi.org/10.1109/ICSMC.2004.1400763