Warning: This Robot is Not What it Seems!

Exploring Expectation Discrepancy Resulting from Robot Design

Lena T. Schramm[1,2], Derek Dufault[1], James E . Young[1]
scle1036@hs-karlsruhe.de, dufaultd@myumanitoba.ca, young@cs.umanitoba.ca

[1]Department of Computer Science
University of Manitoba, Winnipeg, Canada

[2]Department of Business Information Systems
Karlsruhe University of Applied Sciences, Karlsruhe, Germany

ABSTRACT

People are starting to interact with robots in a range of everyday contexts including hospitals, shopping centers, and airports. When faced with a robot, people with little or no prior experience necessarily build expectations based on the robot's superficial appearances and actions, mediated by any potential tangentially related experience (e.g., media depictions). However, the person's constructed expectations (e.g., that a humanoid robot can hold a conversation) does not necessarily relate to actual robot capability, resulting in an expectation discrepancy. This can create disappointment, when the person notices the limited capability, or misplaced trust, if the person believes a robot is more capable than it is. In this paper we present an initial framework for describing and discussing expectation discrepancy.

KEYWORDS

Robot Design, Perception of Robots, Expectation Violation

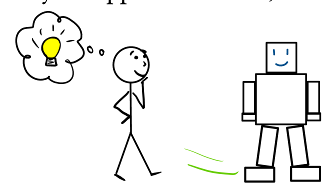
ACM Reference format:

Lena T. Schramm, Derek Dufault, James E. Young. 2020. Warning: This Robot is Not What it Seems! Exploring Expectation Discrepancy Resulting from Robot Design. In *HRI'20 Companion: ACM/IEEE International Conf. on Human-Robot Interaction, March 23–26, 2020, Cambridge, UK.* ACM, New York, NY, USA, 3 pages https://doi.org/10.1145/3371382.3378280

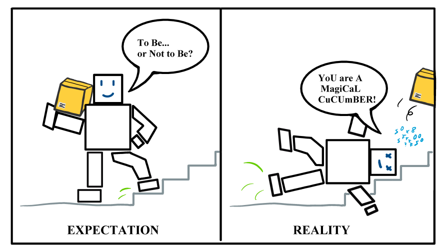
1 Introduction

Social robots are designed with human- or animal-like physical and behavioral characteristics to enhance communication, such as using eyes, facial expressions, or gestures to reflect system state or support collaboration. These techniques are effective in part because people can use their existing social interaction skills to work with the robot; however, as people apply their prior social interaction experience, this may lead to inflated expectations of robot capability. For example, when engaging a robot with a humanoid face that speaks and maintains eye contact, one may intuitively assume that the robot is indeed paying attention to them and can hold a conversation and reason about it. In reality, the robot may simply be following a static script and following faces, without any higher-level reasoning or ability.

This discrepancy between a person's expectations and the reality of a robot's capabilities (Figure 1) poses serious challenges [3]. Once a robot's limited capabilities become clear to a person, often just after short interaction, this creates *expectancy violation* that can result in frustration, disillusionment, and trust reduction [4]. Exploration of expectancy violation has a rich history (e.g., see [2]); researchers in human-robot interaction have compared perceptions from media to interacting with a real robot [1, 6], or how different robot forms can impact expectations [5]. Others have explored the impact of positive (pleasantly surprised) versus negative (disappointed) violations, and how to measure this [4].



a) Person constructs expectations based on observation.



b) Constructed expectations may not match real capabilities.

Figure 1. Discrepancy between expectation based on how a robot appears, versus reality. This paper provides an initial framework for discussing this discrepancy.

On the other hand, if expectations are not violated and people hold inaccurate understanding of robot ability, this could lead to misplaced trust (e.g., as with most Wizard of Oz experiments). Ideally, the community could design robots to leverage the benefits of social techniques (e.g., easy to understand) while reducing the potential issues resulting from expectation discrepancy.

We developed an initial framework, based on in-house workshops and analysis of the field, for describing expectations people develop of robots, and what robot design characteristics impact these expectations. This framework provides a baseline for discussion and further exploring expectancy violation resulting from robot design decisions in social human-robot interaction.

2 Framework Sketch

We identify two distinct components leading to a person's expectations of a robot. One, the robot design itself (appearance, behavior, etc.) indicates potential robot capability; we say the robot *emits* capability signals. Two, based on these signals, a person *constructs* their mental model and thus expectations of the robot (Figure 2). We believe it is important to consider these separately, to understand what robot design characteristics impact expectations, and, what types of sweeping expectations a person may construct based on robot design.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03. https://doi.org/10.1145/3371382.3378280

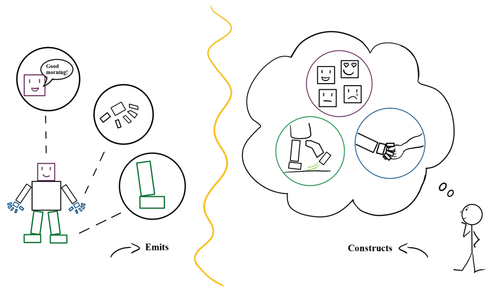


Figure 2. A person *constructs* their mental model based on design characteristics (e.g., shape and form, behavior) a robot *emits*.

2.1 Emission of Potential Capability

By having a particular form or making actions, a robot signifies potential capability. We categorize emissions as follows:

Life-like – People may associate robots with human, animal, or insect-like designs with abilities matched to those designs. This ranges from whole-body form, where one may expect a humanoid to interact socially, to specific body parts, where one may expect a robot with articulated hands to be able to pick up objects or do fine manipulations. Of note are social features, such as a humanoid face, where a robot gaze following or smiling suggests complex underlying attention and emotional structures, even if none exist. This includes both static features (such as having a nose) and dynamic interactions (such as waving at a person, or rhetorical actions such as telling a joke). In short, robot designs inspired from life will likely generate commensurate expectations, even if the robot has no such complex underlying ability.

Consequential – People may assume that a robot with specific elements relating to function has the related capability: one may assume a robot with a camera can see, with an antenna can network, or with tracks can cross uneven terrain, etc. This applies to visible sensors and actuators that people can interpret; people likely recognize a camera lens (and expect it can see), but may not recognize a depth sensor or LIDAR. The robot's form and material construction further suggest capability, for example, a robot with an empty platform or box on top may be expected to carry things, and a metal robot with a severe design may be assumed to be strong and robust. Finally, behavior such as navigating or sorting objects suggests sensing, computational, and planning abilities.

Exposition – How a robot is introduced, how it explains itself, and what task it is used for, can shape expectations. For example, whether a robot is said to be autonomous or remotely controlled, or what keywords are used to explain it (e.g., emotional, artificial intelligence, etc.), will impact what existing understanding people draw from to shape expectations.

2.2 Construction of Expectations

A person observes emitted signals and exposition, and uses prior experience to construct a mental model of robot ability. We note a contrast between minimal information emitted (e.g., looks like has hands) and potential complex mental model construction (e.g., it can do fine manipulations). We break this construction down into two primary categories: mechanical and life-like capability.

We first address mechanical abilities:

Physical Ability – Robot ability in the physical world, including movement, balance, noise, sensing (sight, touch, smell, etc.), and advanced tasks such as writing with a pen, or pushing objects around a room. This also includes robot robustness / fragility.

Computational Ability – People apply understanding of computers to robots, such as the ability to save and retrieve data, perform calculations, logical decisions, or use the internet. People may assume a robot can record video, perform filters and identify people, and post it on a website.

Next we address life-like capabilities:

Non-Social Cognition – A robot that appears to act autonomously may give the impression that it has life-like cognitive abilities, including ability to learn and engage with its environment.

Social Cognition – People may believe that a robot has a social awareness and understanding of others' emotions, the social context of interaction (e.g., a formal situation or awkward discussion), and can employ this information in how it interacts.

Emotional System – Assumptions about a robot having synthetic emotions, and the ability to express them. For example, that a robot that smiles or frowns is expressing an internal happy or sad state.

Social Interaction Abilities – This includes the ability for a robot to appropriately use language such as talking, sign language, gestures, eye gaze in a socially relevant and appropriate manner.

Pseudo-Consciousness – We avoid a philosophical discussion about the nature of artificial consciousness, but note that people's mental models of robot capability may include things typically associated with consciousness. This includes the robot having its own intentions or goals, self-awareness, and creativity.

2.3 From Emission to Construction

Although the connection between robot emissions and a person's construction may seem trivial, our framework highlights how simple emissions (e.g., smiling) can lead to complex constructions (e.g., emotional system, social interaction abilities, and pseudo-consciousness); none of which necessarily reflect robot ability. Designers should carefully consider what constructions their robot designs may lead to, as a starting point for considering the larger question of expectation discrepancy, between what people may assume a robot can do, and what it is capable of.

3 Conclusion

As social robots are designed using humanoid or animal-like forms, and use social communication techniques (faces, gestures, speech) for natural interaction with people, we must consider what expectations people build from these interaction techniques. Thus, ultimately robot designers must consider how which features they embed in their robot, will lead to which expectations, and how this will impact interaction. This larger issue, of expectation discrepancy, can lead to disappointment, misplaced trust, or other critical issues that will ultimately affect robot acceptance.

In this paper, we present an initial framework for describing robot design features that suggest capability, and the kinds of mental models people may construct about these robots. We envision that this will serve as a useful tool to support ongoing research into analyzing expectation discrepancy, and in our own work we have begun investigating how to create measurement tools to help researchers evaluate and better understand what expectations result from a robot design. Ultimately, we envision that this will move the field forward toward having better knowledge and tools, to enable designers to create robots that leverage social interaction techniques, while controlling the resulting expectations.

REFERENCES

[1] Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S. and Tscheligi, M. 2013. The good, the bad, the weird: Audience evaluation of a "real" robot in relation to science fiction and mass media. *ICSR 2013*. 8239 LNAI, (2013), 301–310. DOI:https://doi.org/10.1007/978-3-319-02675-6_30.

[2] Burgoon, J.K., Newton, D.A., Walther, J.B. and Baesler, E.J. 1989. Non-verbal Expectancy Violations. *Journal of Nonverbal Behavior*. 55, 1 (1989), 58–79.

[3] Collins, E., Fischer, K., Malle, B.F., Moon, Aj. and Young, J.E. The Broader Context of Trust in HRI: Discrepancy between Expectations and Capabilities of Autonomous Machines. *Dagstuhl Reports: Ethics and Trust: Principles , Verification and Validation*. M. Fisher, C. List, M. Slavkovik, and A. Weiss, eds. 82–85.

[4] Komatsu, T., Kurosawa, R. and Yamada, S. 2012. How Does the Difference Between Users' Expectations and Perceptions About a Robotic Agent Affect Their Behavior? *International Journal of Social Robotics*. 4, 2 (2012), 109–116. DOI:https://doi.org/10.1007/s12369-011-0122-y.

[5] Kwon, M., Jung, M.F. and Knepper, R.A. 2016. Human expectations of social robots. *ACM/IEEE International Conference on Human-Robot Interaction*. 2016–April, (2016), 463–464. DOI:https://doi.org/10.1109/HRI.2016.7451807.

[6] Sandoval, E.B., Mubin, O. and Obaid, M. 2014. Human Robot Interaction and Fiction: A Contradiction. *ICSR. Lecture Notes in Computer Science, vol 8755*. 54–63.