An Analytical Framework to Examine and Describe People's Expectations of Robots

by

James Matthew Berzuk

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

Department of Computer Science
The University of Manitoba
Winnipeg, Manitoba, Canada

Thesis Advisor
James E. Young

Author
James M. Berzuk

An Analytical Framework to Examine and Describe People's Expectations of Robots

Abstract

We engaged with the problem of expectation discrepancy in human-robot interaction: a known challenge in which the expectations people form when interacting with a social robot may not align with its actual capabilities. This misalignment, an expectation discrepancy, can disappoint users and hinder interaction. While research has proposed ways to mitigate expectation discrepancy, designers lack a systematic approach to analyzing and describing expectations people form of their robot. A more rigorous theoretical framework is a necessary step towards designing robots to purposefully engineering desired expectations. We consulted theories and models from psychology and sociology on expectations between people, and conducted a survey of expectations in human-robot interactions. Through this we developed an analytical framework consisting of a novel model of the cognitive process of human-robot expectation formation, as well as a taxonomy for classifying the types of expectations they form. We finally propose preliminary methods for designers to use this framework as a tool to support systematic analysis of how and why people form expectations of a given robot and what those expectations may be. Such understanding can empower designers with greater control over people's expectations, enabling them to combat problems of expectation discrepancy.

Acknowledgements

Firstly, I would like to thank my advisor, Dr. James Young. Your immense support and guidance has been absolutely invaluable, teaching me how to conduct research truly from the ground up. I am so grateful for all the opportunities you arranged, especially joining you in Japan. Thanks to you and your family for making me feel so welcome when I was there.

Thank you as well to my committee members, Dr. Andrea Bunt and Dr. David Gerhard, for all of the very helpful feedback and interesting discussions we have had over this project.

I would also like to thank all the students of the HCI Lab. I am so grateful for all the support you have provided, starting from when I joined the lab remotely during the pandemic and knew very little about research, through to all the fun memories we have had at lab game nights and other events. I would especially like to thank Raquel and Danika for welcoming me into your studies and allowing me to partake in other sides of research I was not able to experience with my own project. Thank you also to the Hokkaido HCI Lab, especially Yuki and Kan, for welcoming me into their lab and sharing so many fun experiences.

Lastly, I would like to thank my family for all their love and support. Mom, thank you for inspiring me with your own academic career, and for the many hours helping proofread this thesis. Dad, thank you for all the encouragement, helping me to see the value in my work even when I was struggling. Cass, thanks for helping me brainstorm so many ideas. And finally, to my partner Lauren, I cannot thank you enough for being there for me through all of this, listening through all the challenges and celebrating all the successes.

Table of Contents

List of Figures

List of Tables

Publications

Some ideas in this thesis have appeared previously in the following publication:

> Berzuk, J. M., & Young, J. E. (2023). Clarifying Social Robot Expectation Discrepancy: Developing a Framework for Understanding How Users Form Expectations of Social Robots. *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 231–233. https://doi.org/10.1145/3568294.3580078

In addition, much of Section 2.1 is drawn from the following research paper in preparation:

> Berzuk, J. M., Corcoran, L., Szilagyi, K., & Young, J. E. Knowledge Isn't Power: The Ethics of Social Robots and the Difficulty of Informed Consent. To be submitted to the *International Journal on Social Robotics.* Manuscript in preparation.

Chapter 1 Introduction

The field of human-robot interaction examines how people interact with robots of many different varieties and aims to support the design of robots that can achieve smoother and more successful interactions. Within this field, one area of particular focus is social robots. Social robots are designed to simplify collocated interaction with people by leveraging life-like social features that people can readily understand (Breazeal, 2003).

When a person interacts with a social robot, they may form a plethora of expectations of the robot based on its design and their initial expectations. For example, a person may reasonably assume that if the robot has hands and fingers, then it can pick up items (Schramm et al., 2020; e.g., Figure 1). However, the robot may not have this capability, creating an *expectation discrepancy* (Schramm et al., 2020) where people may not only misunderstand

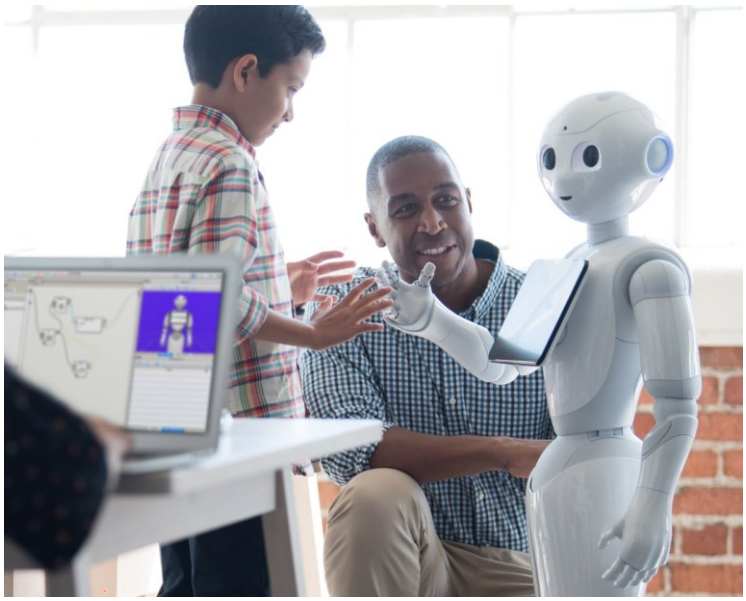

Figure 1: The Softbank Pepper robot (SoftBank Robotics America, Inc., n.d.) has mobile, human-like hands, which give the incorrect impression that it can pick up items and manipulate objects in a human-like way.

how to interact with the robot, but may be surprised and disappointed by a lack of ability, impacting the quality and success of the interaction (Komatsu et al., 2012). These misunderstandings can have far-reaching implications including misplaced trust and a host of impacts on how robots integrate into society (Sharkey & Sharkey, 2021), placing the issue of expectation discrepancy – and managing it – at the center of successful human-robot interaction. While some work has investigated moderating user expectations (e.g., Kwon et al., 2018; Paepcke & Takayama, 2010), the field does not yet have a systematic approach to understanding and analyzing users' expectations.

We have developed an analytical framework that designers can use to support examination and explanation of potential or observed expectations of their robots. Our framework consists of two key components that may serve as tools for designers and researchers: a model of the cognitive process by which people form expectations, and a taxonomy for organizing and describing these expectations. Further, we devised preliminary applied techniques for leveraging our framework to analyze expectations of robots in practice. We use case studies to demonstrate this framework and highlight its applicability to real robots, and conclude with a critical reflection on the framework to identify its strengths and limitations.

1.1 Expectations of Social Robots

The expectations people form of robots are complex, and require deeper consideration than simply referring to them vaguely under the heading of 'expectations'. They can emerge from a range of sources, including decades of fanciful media depictions (Bruckenberger et al., 2013; Sandoval et al., 2014), and are heavily influenced by the robot's own design (Kwon

et al., 2016; Natarajan & Gombolay, 2020). Robot designs can align a robot with some known mental category (e.g., an animal) and thus can imply capabilities which are commonly associated with that category (e.g., can think, has an emotional system, etc.; Cross & Ramsey, 2021). This allows people to leverage their existing knowledge when encountering a social robot for the first time. Outwardly human- or animal-like design features can also promote familiarity (Breazeal, 2003) and leverage empathy in users (Riek et al., 2009). This complex interaction, involving prior depictions of robots and similarities to familiar entities, motivates more work on explaining how people form expectations of robots, and what types of expectations they form.

1.1.1 Expectation Discrepancy

Life-like design features may be effective for certain interaction goals, but they may simultaneously lead to inflated associated expectations of human- or animal-like capability. When a robot is, often inevitably, incapable of meeting these inflated expectations, this produces an *expectation discrepancy* (Kwon et al., 2016; Schramm et al., 2020). This discrepancy disrupts the interaction as the user attempts to map their existing knowledge onto the robot but finds it ineffective, and can disappoint the user and hinder the interaction (Komatsu et al., 2012).

Expectations may emerge in part from robot design choices. As such, we may be able to mitigate or even avoid inflated expectations, and expectation discrepancy, by designing robots in ways that more accurately imply their capabilities (Kwon et al., 2018; Paepcke & Takayama, 2010) However, there is as of yet no systematic approach for doing so. A step

toward this goal that we take in this work is to understand how and why people form expectations of the robots they encounter, and what kinds of expectations they will form.

1.1.2 Defining Expectations

It is important to improve clarity of meaning and scope surrounding the term 'expectations' in the context of human-robot interaction. While some works on human-robot expectations are careful to employ a precise definition (e.g. Rosén et al., 2022 defines expectations as "believed probabilities of the future"), others use the term without any explicit definition, often using the term less formally to refer to impressions of a robot's capabilities (e.g. Kwon et al., 2016; Schramm et al., 2020). In absence of any universally-adopted definition, we note that this looser application of the term is commonly found in literature on expectation discrepancy, which concerns the difference between real and expected capability (Komatsu et al., 2012; Kwon et al., 2016; Schramm et al., 2020). As such, in this work, we use the term *expectations* to refer to a person's beliefs, conscious or otherwise, about a robot's capabilities and behaviour. Similarly, we use the term *expectation discrepancy* to refer to any disparities between a person's expectations of a robot and the robot's true capabilities and behaviour.

1.2 Research Questions

We formalize our investigation with four key research questions:

RQ1: How is the research community engaging with the concept of human-robot expectations?

We must first understand how the human-robot interaction research community is currently engaging with the concept of human-robot expectations, and especially with the problem of expectation discrepancy. In particular, we will look to determine whether the community has a consistent approach to these topics. This will allow us to highlight areas in need of unified approaches and vocabulary, as well as to work within the prevailing tradition where consistency is found.

RQ2: What is the process by which people form expectations of robots they encounter?

If we are to influence, in a systematic manner, a person's expectations of a robot, we must first understand how these expectations are developed. We aim to describe this formation process. This includes identifying the inputs into the process (that is, the factors that determine resulting expectations, be they properties of the person, the robot, or the context of the interaction), as well as describing its composition.

RQ3: What are the patterns in expectations that people form of robots, and can we distill them into a taxonomy?

In order to engage with instances of expectation discrepancy, we require a structured and unified approach to describing expectations. It is necessary to identify common patterns in people's expectations of robots in order to group and classify them, which will allow us to discuss them at a higher level, rather than focusing on individual, disparate expectations.

RQ4: How can our improved knowledge of human-robot expectations be used by robot researchers and designers to examine and explain expectations of their robots?

Using our understanding of how expectations of robots are formed and of how they can be described and classified, we aim to design practical tools and techniques that allow robot designers to leverage this understanding toward combatting the problem of expectation discrepancy. We must demonstrate how the theoretical understanding developed in this work can be used to analyze and explain users' expectations and ultimately support designers in influencing those expectations to enable more successful interactions.

1.3 Methodology

To engage with our research questions, we conducted two major investigations: we reviewed research from psychology and sociology on expectations between humans to develop a model of the process by which people develop expectations of robots, and we conducted a broad informal survey of expectations in human-robot interaction literature to develop a taxonomy for classifying expectations.

1.3.1 Synthesis of Theoretical Literature

We developed our understanding of how people form expectations of robots by first consulting how people form expectations of other people. We analyzed prominent theories from psychology and sociology that describe how people form and manage expectations of each other (human-human expectations), synthesizing them from the perspective of interaction with robots. This synthesis resulted in a novel model of the cognitive process of human-robot expectation.

1.3.2 Survey of Expectations

To understand the patterns that exist across user expectations and inform how we can describe and classify them, we conducted a broad, informal survey of existing robots, prototypes, behaviors, and literature on expectations of robots. We analyzed this collection for potential expectations to assemble an initial expectations corpus. We then conducted a thematic analysis on this corpus, identifying commonalities and salient patterns, resulting in a novel two-dimensional taxonomy of human expectations of robots.

1.3.3 Evaluation

Given the theoretical nature of our work, the background literature and theories, as well as the logic and validity of our synthesis, serve as the primary measure of the validity of the work. To further illustrate how our framework can be leveraged in practice to support designers and researchers, we developed two preliminary methods of applying our tools to real designs, which we demonstrate using case studies. Finally, we conduct a critical evaluation of our work, looking both to our background literature as well as our experiences with applying the theoretical framework in our case studies, to identify the effectiveness and limitations of our framework and to highlight opportunities for future work.

1.4 Contributions

The process described in Section 1.3 resulted in two analytical tools: a model of the cognitive process by which people form expectations of robots, and a two-dimensional taxonomy for classifying the expectations they form. Robot designers can use these tools for in-depth analysis and exploration of expectations.

Our model of the cognitive process of human-robot expectation formation consists of an enumeration of the influencing factors (e.g., robot design, personal experience, interaction context) and stages that a user goes through to develop, maintain, and update their expectations. Grounded in literature on expectations between people, this model offers a theoretically-backed understanding to explain why a person may form a particular expectation.

Our taxonomy of human expectations of robots consists of two dimensions: *domain of expected capability* and *level of expectation abstraction*, which together can be used to describe and organize expectations according to common patterns observed in our survey of human-robot expectations literature. This includes a visualization of the classification space that can provide a graphical representation of a user expectations and discrepancies.

We further developed two preliminary analytical techniques for applying these tools to real robots: *systematic expectation dissection* and *cognitive expectation walkthroughs*. These techniques provide an initial guide toward employing our tools in the real world, though we envision they will need to be refined through practice in the field.

Altogether, these tools compose an analytical framework, together with preliminary techniques for practical application, to support designers in examining and explaining expectations of their robots, providing a foundation for improving designer control over user expectations in order to mitigate expectation discrepancy and achieve more successful human-robot interactions.

1.5 Thesis Overview

In this chapter, we outlined the problem of expectation discrepancy and our strategy to develop an analytical framework to engage with it. In the following chapters, we will address each of our four research questions in order.

In Chapter 2, we review prior work on expectations in human-robot interaction, as well as on frameworks and other analytical tools in human-robot interaction more broadly, to understand how the field is currently engaging with human-robot expectations and expectation discrepancy (RQ1) and further inform our approach.

In Chapter 3, we explore literature from psychology and sociology on expectations between people and synthesize it from the perspective of human-robot interaction to develop a model of the cognitive process by which people form expectations of robots (RQ2).

In Chapter 4, we conduct a broad, informal survey of expectations in human-robot interaction literature to build a corpus of potential expectations, which we analyze to develop a taxonomy of human expectations of robots (RQ3).

In Chapter 5, we present two preliminary analytical techniques that designers can use to apply our process model and taxonomy and examine and explain users' expectations of their robots (RQ4). We further demonstrate these techniques using case studies.

In Chapter 6, we conduct a critical reflection on our framework, evaluating its utility to designers and researchers, as well as its limitations both theoretically and in practical application.

Finally, in Chapter 7, we conclude with a summary of our work, recalling our research questions and highlighting how our analytical framework may be employed toward mitigating expectation discrepancy. Additionally, we recall our framework's limitations and offer recommendations for its successful application to real robots, as well as opportunities for future research.

Chapter 2 Background and Related Work

Before we develop a formal understanding of expectations in human-robot interaction, it is essential that we ground ourselves in how the field is currently engaging with this subject. In this chapter, we will answer RQ1: *How is the research community engaging with the concept of human-robot expectations?*

We begin this chapter by reviewing what social robots are and what makes interaction with them distinct from interaction with other entities, both technological and living. In doing so, we highlight the need to consider expectations of robots independently from expectations in other, more familiar interactions.

Following, we review the wide body of research on how a robot's design can influence a person's impression of it, as well as on the interaction between them. These works highlight the many ways in which a robot can impact user expectations, and offer perspective on how the field is discussing those expectations.

We will then look specifically at literature on expectation discrepancy in human-robot interactions. This serves to address the core of our research question: we will examine how expectation discrepancy is being framed and what strategies have been employed to counter it. Through this, we will identify areas where our work can support a more systematic approach to understanding and mitigating expectation discrepancy.

Finally, we inform our approach by examining the extensive use of frameworks in the field of human-robot interaction to support engagement with particular ideas and challenges.

2.1 Properties of Social Robots

Note: Much of this section is drawn from the following research paper in preparation:

Berzuk, J. M., Corcoran, L., Szilagyi, K., & Young, J. E. Knowledge Isn't Power: The Ethics of Social Robots and the Difficulty of Informed Consent. To be submitted to the *International Journal on Social Robotics.* Manuscript in preparation.

Interaction with social robots is a unique social phenomenon, different from interacting with other technologies or living entities. This differentiation stems from the intersection of robots' social and physical embodiment, which can resemble humans or other living things, with the superhuman capabilities of a computer. This positions social robots as a novel kind of interaction entity. Robots act like they are alive, while remaining a technological artifact. This deceptive pattern can distort expectations, encouraging people to map their expectations of more familiar entities, both living and technological, onto robots which may not cleanly reflect either. Understanding this unique positioning is critical to understanding how people's expectations of robots may differ from their expectations of both living beings and machines.

2.1.1 Designed Sociality

Social robots are explicitly designed to engage with humans' emotions and social instincts. For example, they can be designed with outwardly human- or animal-like features (e.g., humanoid shape, realistic voice, life-like gaze-following, etc.; Breazeal, 2003; Phillips et al., 2018) and exhibit displays of emotion in order to facilitate interaction or achieve a related goal. This is much less common with more conventional technologies that people are

familiar with, which are typically designed for more mechanical or direct informational interaction.

When a person encounters a social robot, the life-like features encourage anthropomorphism or zoomorphism (more generally, *animorphism*), where the person attributes the robot with life-like (e.g., human- or animal-like) characteristics to help them understand it and determine how to interact with it (Epley et al., 2007; Złotowski et al., 2018). This process can vary, from serving as a social expedient leveraging existing knowledge to facilitate interaction, to regarding the robot as a social peer and using it to fulfill social needs (Epley et al., 2007). As this process may be grounded in an evolved human tendency (Złotowski et al., 2015), it may be quite difficult for individuals to overcome.

Robots presenting life-like abilities and features can be viewed as a form of deception that may lead users to make incorrect assumptions about a robot's abilities (Sharkey & Sharkey, 2021). Despite the lack of genuine substance behind these social interactions, they can have real impacts on users. For example, one social robot pressured individuals into assisting it for nearly 30% longer by utilizing a script that leveraged the user's cultural background in a socially intelligent manner (Sanoubari et al., 2019), while another social robot persuaded people to disclose intimate information by first divulging its own 'secrets' (Y. Moon, 2000). Robots have also been shown to be able to guide people toward poor decisions, such as pouring orange juice over a plant (Salem et al., 2015), or taking a wrong turn during an evacuation scenario (Robinette et al., 2016; Figure 2). These striking reactions highlight the power of deceptive, animorphic robot designs.

Figure 2: This robot was able to guide participants through a false exit in a mock evacuation scenario despite the sign for the real exit being in direct view (Robinette et al., 2016).

Relatedly, interaction with social robots over time can lead to the development of parasocial relationships (Noor et al., 2021): unilateral emotional connections that people can form with artificial entities, fictional characters, and media personalities (Brown, 2015). These interactions create an illusion of reciprocity between the person and robot, in which the latter's behaviour encourages the fiction of a mutual interpersonal connection. These connections, though illusory, can have real tangible effects on people. The effects of parasociality may be positive, such as improving a person's perceived sense of well-being (Noor et al., 2021), or potentially hazardous, such as influencing a person's spending habits, by promoting a particular purchase (Hwang & Zhang, 2018) or stimulating impulse buying behaviour (Zafar et al., 2020). While such effects may be expected when dealing with other people, this influx of sociality into traditionally non-social machines allows for social manipulations which may defy a person's expectations of the interaction.

2.1.2 Robot Physicality

A defining trait of social robots is their physicality (Hegel et al., 2009), as social robots dynamically occupy space in the user's environment. Often capable of moving autonomously, social robots garner a heightened sense of presence and attention, explained in part by human instincts toward detecting motion to identify others and assess threats (Simion et al., 2011); even from infancy, humans have been observed to view life-like patterns of movement as indicative of a social agent. For example, infants respond to robots as social agents, following their gaze as they would a person (Meltzoff et al., 2010). These anthropomorphizing reactions—to ascribe human motivations to non-human objects—also happen when observing moving objects with no otherwise-life-like characteristics. For example, people assign agency and emotional state to the motion of a collection of abstract shapes (Heider & Simmel, 1944; Figure 3), and even to a stick that simply moves in regular or irregular patterns (Harris & Sharlin, 2011). While the exact mechanisms behind this

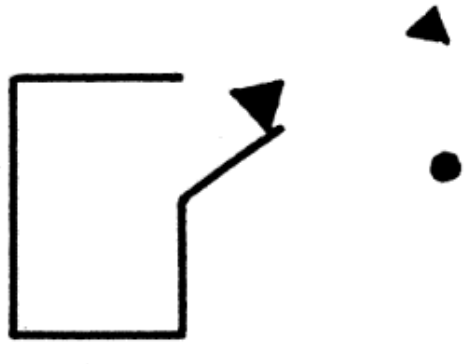

Figure 3: Simple abstract shapes, once animated, were sufficient for people to develop entire narratives about their decisions and emotions to explain their movements (Heider & Simmel, 1944).

phenomenon are not well understood (Cross & Ramsey, 2021), this evidence suggests that the effect is a natural occurrence, grounded in biology.

While virtual on-screen agents also garner strong reactions (e.g., as in popular media), there is a large body of research that highlights stark differences between how people respond to robots in comparison with more traditional media (Hegel et al., 2009). For example, studies have demonstrated that participants assign greater social presence to physically present robots (Jung & Lee, 2004) and provide them with more "personal space" (Bainbridge et al., 2011) versus on-screen virtual agents on a monitor, and rate robots more amicably (Jung & Lee, 2004) in general. Further, people can be more compliant with the requests of physical robots in comparison to virtual agents (Bainbridge et al., 2011). Social robots have the potential to elicit stronger emotional responses than virtual agents; participants express more sympathy for a robot than a virtual agent that was placed in a distressing situation (Seo et al., 2015; Figure 4). People even have heightened emotional responses to a co-present robot versus a live video feed of a remote robot (Li, 2015), an overall effect which may be linked to the greater perceived physical size of collocated robots in comparison to those displayed on a smaller screen, similar to how people respond to taller individuals versus shorter ones (Li, 2015).

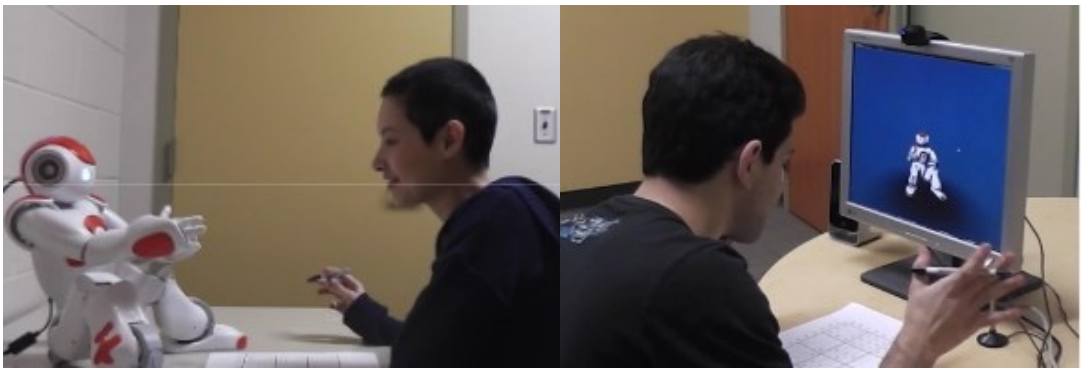

Figure 4: The physical robot (left) was able to elicit more empathy in participants than the virtual equivalent (right) (Seo et al., 2015).

2.1.3 Superhuman Abilities

Although social robots' design may present them as life-like entities, their computation and networking power gives them a wealth of superhuman capabilities not outwardly reflected in their designs. This includes the ability to collect incredible amounts of data from sensors or networked sources and process it at incredible speeds, using this to adjust social interactions in real time. While a person might expect a salesperson to leverage surface observations about them to adjust their sales pitch in real time, a sales robot could leverage high-speed cameras and advanced algorithms to monitor facial and body motions, inferring the human's state, while simultaneously analyzing the person's entire public social networking record, to tailor an efficient and personalized pitch. All of this is invisible to the user as there is no outward indication of the robots' internal machinations.

Robots are not bound by the limitations that may be suggested by their designs; for example, while a robot may close its eyes to signal that it is not watching, it can still observe from cameras not in the eyes (Kaminski et al., 2016). Concurrently, robots can be purposefully designed with constraints and operational inflexibilities to limit and steer user actions, as

a kind of "dark pattern" (Dula et al., 2023); for example, a robot may pass something to a person but not have the physical capacity to take it back, forcing the user to keep it. These limitations can force a user into making choices they would not otherwise make when interacting with a person.

In human-human interactions, there is an expectation that both parties will be similarly vulnerable, engaged in principles of reciprocity that contribute to a feeling of mutual trust (Y. Moon, 2000). Social robots seek to influence humans but are not limited by corresponding human social features that would cause them to be influenced in return. This violates basic social expectations that are brought to interactions with other people. By allowing an illusion of reciprocity to persist, in which a person may believe that their actions are affecting the robot similarly to how they would a human, social robots betray that trust and mislead users about the true nature of the interaction, compromising the users' ability to make informed decisions. Social robots are not necessarily constrained by the rules and assumptions of human-human relationships (de Graaf, 2016).

2.1.4 Social Robots Are Unique

Taken together, the active design choices directed towards robots' sociality leverage human biological instincts to demand attention and generate emotional response. These phenomena are particularly pronounced with physical robots as compared against virtual agents. Because of their physical presence and mobility, social robots can manipulate users' internal emotional states, establishing long-term feelings of connection and obligation. At the same time, robots are different from actual living social agents in their ability to draw on

computational power and massive quantities of data, and their inability to meet basic social expectations of reciprocal influence and mutual vulnerability.

These properties together both relate social robots to other living and technological entities, and ultimately distinguish them. When interacting with a robot, people may apply their expectations of those more familiar entities, but this unique combination may influence the process in novel ways and become a source of expectation discrepancy. It is for this reason that specific consideration is necessary when seeking to understand human-robot expectations. Thus, in our work we combine theoretical understandings of expectation formation between people with the existing knowledge of expectations in human-robot interaction to develop a framework suited to this unique intersection.

2.2 Robot Design and Expectations

The impact of a robot's design on expectations, and thus interaction, is well documented, with a large body of work exploring the impact of specific robot design factors. In this section we will first review works that consider the impact of robot form and aesthetics, and then those that consider the impact of robot behaviour.

2.2.1 Robot Form

Much of this work considers effects of a robot's aesthetic form, following a common pattern where participants are shown a series of robot variants and asked to rate them on specific metrics. For example, Rosenthal-von der Pütten & Krämer (2014) presented participants with pictures of 40 different robots and had them evaluate each according to 16 axes, such as likability and familiarity, in order to understand how life-like robot designs contribute

to feelings of uneasiness (known as the *uncanny valley*). Schaefer et al., 2012 employed a similar approach to explain how aesthetic form can impact perceived trustworthiness.

One common focus is in linking features to anthropomorphism (Phillips et al., 2018) and to how this impacts user reactions. For example, Fortunati et al. (2023) examined how a resemblance to humans impacts perception of cognitive ability (Figure 5). Haring et al. (2013) found that trust in safety of a human-like robot increased after interacting with it, while Natarajan & Gombolay (2020) found that perceived anthropomorphism contributes to trust more generally.

Another property of a robot that can impact expectations is the sounds that it produces: adding mechanical-sounding noises to an otherwise identical movement can make the movement appear less controlled and precise (Robinson et al., 2021), while emitting barely-audible, low-frequency infrasound can make a robot's communication appear happier (Thiessen et al., 2019). Comparisons have also been made between virtually-embodied agents and physically-embodied with robots, with people exhibiting greater empathy for physical robots (Seo et al., 2015), but perhaps similar (van Maris et al., 2017) or even reduced levels of trust (Reig et al., 2019).

Figure 5: Fortunati et al. (2023) compared perceptions of cognitive ability across these four robots with differing degrees of resemblance to humans.

More holistically, Dennler et al. (2023) explored using metaphors to explain and understand robots, where placing a robot into a known social category can support a person to understand a robot, and shape expectations, in relation to a familiar entity (Figure 6).

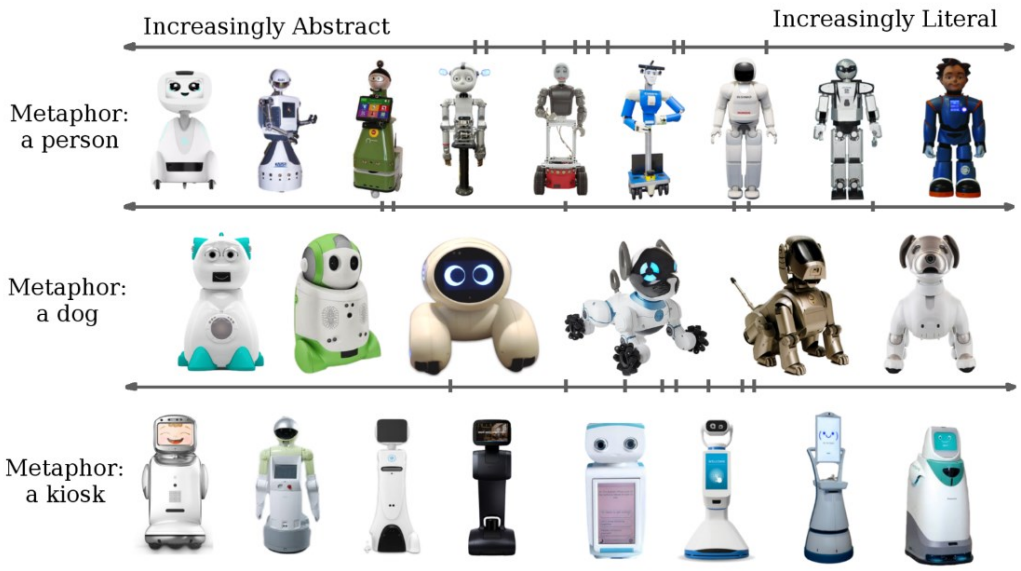

Figure 6: Dennler et al. (2023) organized robot designs in terms of metaphors to more familiar entities in everyday life.

2.2.2 Robot Behaviour

Similar to those on robot forms, many works have examined the impact of robot behaviors on user expectations. Within this category, there are numerous examples that focus on the use of social cues to build trust and social presence (K. Xu et al., 2023). Employing happy and fearful facial expressions has enhanced participant's impressions of a robot (Eyssel et al., 2010). Combining facial expressions with expressive verbal communication has been shown to increase likability, even when the robot is less efficient at its tasks (Hamacher et al., 2016; Figure 7). Stanton & Stevens (2017) experimented with robots maintaining eye contact with participants, and found that perception of such staring was gender-mediated, with excessive staring degrading trust from female participants. The robot's 'gender' has also been considered; Bryant et al. (2020) tested to see if a robot expressing its gender as male, female, or neither impacted participants' perceptions of its competency, but found no significant difference. These works demonstrate the complex, often opaque relationship between a robot's designed conduct and the way that people respond to it.

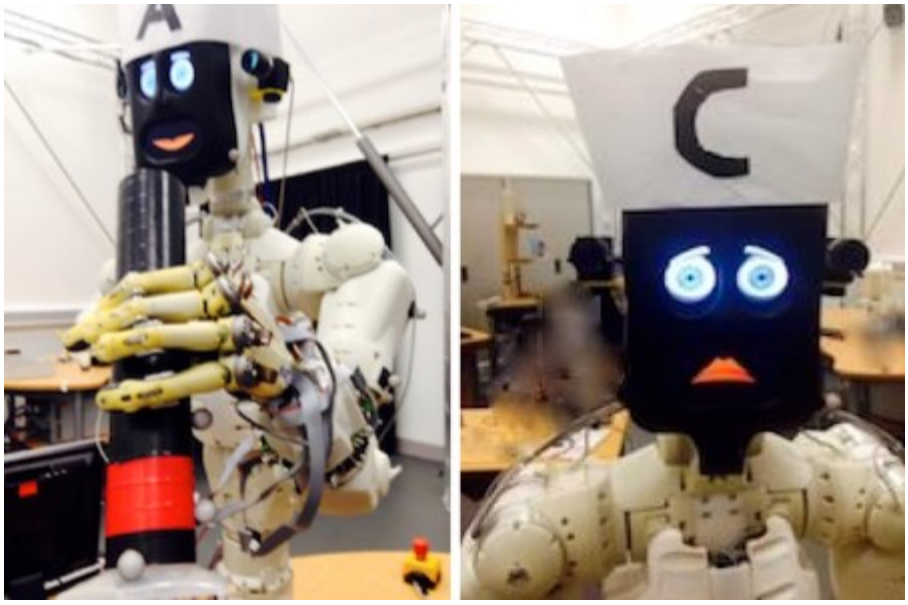

Figure 7: The BERT2 platform utilizing facial expressions for expressive communication to enhance likability with participants (Hamacher et al., 2016).

Another focus of research on robot behaviour is to test the effects of robots making mistakes on perceptions and interaction. Mirnig et al. (2017), for example, found that a robot making mistakes while instructing participants on a task made it more likeable, and found no significant impact on the perceived intelligence of the robot. This positive effect of mistakes is well-attested: multiple studies have shown that robots cooperating with participants are regarded more positively when they make mistakes, even when those mistakes come at the expense of the participant's performance in their shared task (Ragni et al., 2016; Salem et al., 2013). Mistakes are not an unambiguous positive however; speech errors can make a robot appear more familiar but less sincere (Gompei & Umemuro, 2015), while giving faulty instructions can degrade trust and perceived reliability (Salem et al., 2015). Once again, we find the impact of a robot's design on expectations to be nuanced and complex.

2.2.3 Providing a Holistic Perspective

Our research complements this growing body of largely-empirical work that outlines precisely how specific robot designs can impact expectations, by providing encompassing theoretical tools for analyzing and exploring the observed effects. Through our grounding in literature on expectations between people, we offer a procedural, explanatory perspective on how metaphor and resemblance to known entities can contribute to a person's expectations of a robot.

2.3 Expectation Discrepancy

A range of work has outlined impacts of robot expectation discrepancies – where people construct expectations that do not match actual abilities (Kwon et al., 2016; Schramm et al., 2020; Figure 8). This often highlights user disappointment, such as when a person attempts to talk with a robot that cannot converse (de Graaf et al., 2015). These discrepancies can detract from a user's experience (Lohse, 2011) and create a sense of incompetence and lower trust (Salem et al., 2015).

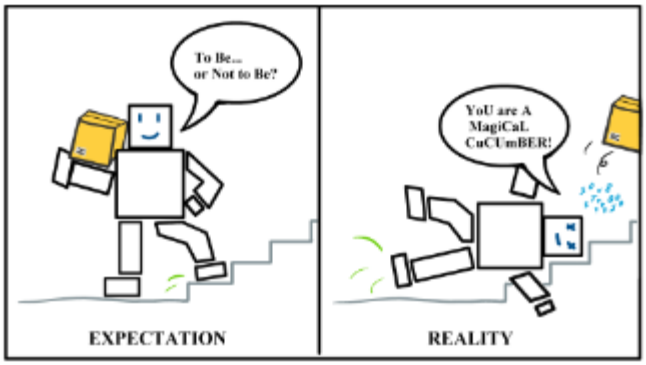

Figure 8: Schramm et al. (2020) depicts the disparity between the advanced, human-like conception of robots often portrayed in media and the technical challenges found in many robots today.

The effects of expectation discrepancy can be more nuanced, however. As discussed in Section 2.2.2, robot failures are not always simply negative, and can enhance familiarity and likeability (Gompei & Umemuro, 2015; Mirnig et al., 2017; Ragni et al., 2016; Salem et al., 2013). Alternatively, a robot exceeding expectations may cause a person to trust and rely on it more (Komatsu et al., 2012), though if a person does not notice a robot's lack of real ability, expectation discrepancy can ultimately lead to misplaced overtrust with potentially dangerous results (Sharkey & Sharkey, 2021).

2.3.1 Moderating Discrepancy

Some work investigates ways to moderate user expectations (e.g., to be in line with robot abilities), such as using exposition about the robot's capabilities (Paepcke & Takayama, 2010), aesthetic forms more congruent to function (Collins et al., 2015; Goetz et al., 2003), or the robot itself using expressive gestures of incapability (Kwon et al., 2018).

Our structured tools build on this work by providing a comprehensive basis for designers to systematically analyze designs for potential discrepancies between user expectations and their robots' abilities, and to describe and explain the expectations observed.

2.3.2 Social Robot Expectation Gap Evaluation Framework

Rosén et al. (2022) offers a framework for evaluating expectation discrepancy in users interacting with robots. They adapted a model of expectation formation between people (Olson et al., 1996) to use with robots (Figure 9), and identified a set of factors and

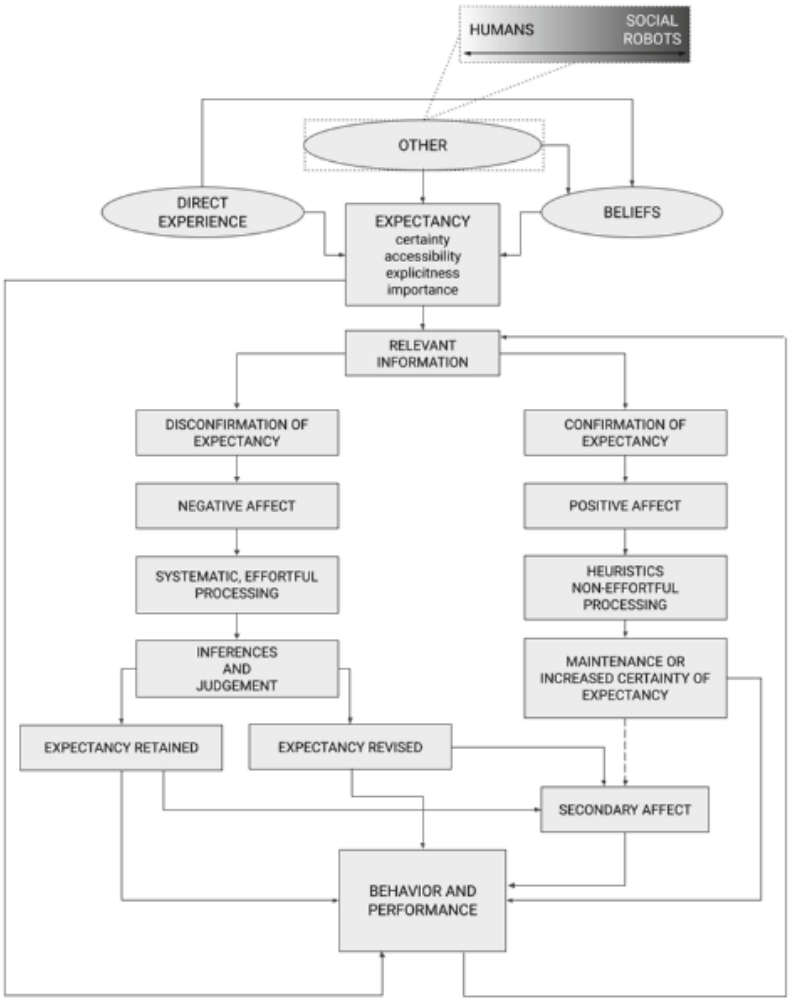

Figure 9: Rosén et al. (2022) modified Olson et al. (1996)'s model of the expectation process for application to human-robot interaction.

corresponding metrics that can be used to measure a person's expectations of a robot, thus enabling evaluation of the level of expectation discrepancy a user experiences with a particular robot. Specifically, the framework measures a user's affect toward the robot, the cognitive load on the user during the interaction, and the degree to which the user expects an easy and pleasant interaction with the robot, and uses these factors as the basis for identifying expectation discrepancies.

Complementary to this focus on interaction outcomes, our work extends this by supporting understanding of the causes of expectation discrepancies, with a particular focus on how expectations evolve through mental simulation and refinement. Further, our taxonomy provides a way to classify expectations and discrepancies according to their content (i.e., what the user expects), such that our cognitive process and taxonomy provide tools for analyzing and explaining discrepancies which may be revealed through the Social Robot Expectation Gap Evaluation Framework (Rosén et al., 2022).

2.4 Frameworks in Human-Robot Interaction

Within the field of human-robot interaction, many different conceptual frameworks have been developed to support researchers and designers in understanding and engaging with challenging topics. At a high level, such frameworks can be divided into those which examine human-robot interactions at a general level, and those which adopt the lens of a particular domain.

There are many frameworks that describe interactions between humans and robots. Kahn et al. (2008) takes a component-focused view of human-robot interaction by compiling a list of frequently-observed patterns in interactions, such as the necessity to recover from mistakes, or to navigate turn-taking in a social activity. Yanco & Drury (2004) offers at taxonomy for classifying interactions according to, among other things, the structural relationships of the participants and their roles in the interaction. Bartneck & Forlizzi (2004) employed a five-dimensional framework for concisely categorizing and contrasting social robots according to their form, interaction modalities, adherence to social norms, autonomy

and interactivity (Figure 10). These frameworks are all quite general, and applicable to most instances of human-robot interaction, but do not offer specific insights for regarding expectations.

Other frameworks target specific domains in order to support engagement with a particular problem. One area of focus is on human-robot dialogue, with frameworks for example classifying dialogue instances as linear or branching in nature (Berzuk & Young, 2022; Figure 11), or identifying technical patterns in dialogue interaction design such as checks for repetition and randomized variation (Glas et al., 2016). Some frameworks relate to particular outcomes, such as considering factors that lead people to accept a robot into their homes (Young et al., 2009), and many consult peripheral areas to incorporate novel perspectives into the field (e.g., consulting literary analysis for human-robot dialogue systems; Berzuk & Young, 2022). Rosén et al. (2022) (discussed in detail in Section 2.3.2) applied this approach

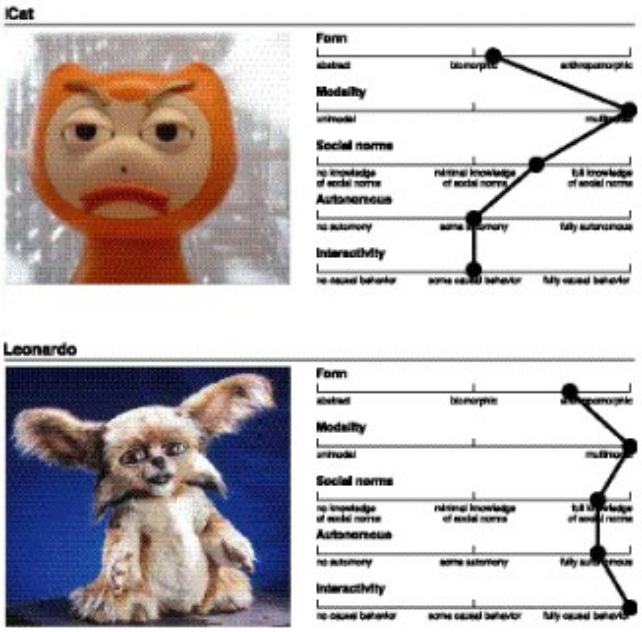

Figure 10: Bartneck & Forlizzi (2004)'s framework can be used to concisely classify different social robots and compare them at a glance.

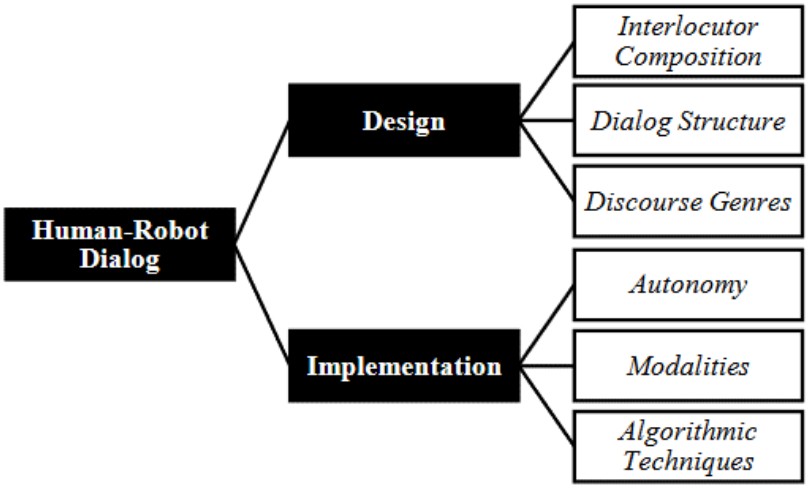

Figure 11: Berzuk & Young (2022)'s framework for describing human-robot dialogue designs identifies key dimensions differentiating various human-robot dialogue interactions and offers a vocabulary for discussing and contrasting them.

to human-robot expectations, developing a framework that can be used to measure expectation discrepancy.

Our work builds on this rich methodological tradition, of synthesizing work from other fields into a framework to provide structure and support analysis of human-robot expectation formation and discrepancy.

2.5 Chapter Summary

At the beginning of this chapter, we asked how the research community is engaging with human-robot expectations (RQ1). We have now explored the existing literature on this subject and have found the community stands to benefit from more consistent perspectives and vocabulary on the subject. While substantial efforts have been made to draw attention to and in some cases identify and mitigate human-robot expectation discrepancy, these efforts can be complemented by an overarching analytical framework for understanding the

issue, as has been developed for other problems in human-robot interaction. Our framework can support designers with theoretically-backed analytical tools to examine and explain expectation discrepancy with their designs in a consistent and systematic manner.

Chapter 3 How People Form Expectations of Robots

Our first step to developing a systematic understanding of people's expectations of robots is to understand how they form those expectations. In this chapter we will engage RQ2: *What is the process by which people form expectations of robots they encounter?*

To engage with this question, we begin by analyzing current knowledge of how people build expectations of their world and other people they encounter, linking these existing theories and ideas to interaction with social robots. We then synthesize these theories from the perspective of human-robot interaction to draw key points for understanding how people form expectations of robots. Finally, we use these points to build a model of the cognitive process of human-robot expectation formation.

3.1 Expectations Between People Explain Expectations of Robots

In this chapter, we will translate how people build expectations of *other people* to understand how they may build expectations of *robots*, despite the fact that robots are not people. We rely on the assumption that people tend to treat physically embodied robots as if they were alive (Złotowski et al., 2018). As discussed in Section 2.1.1, this generally follows the concepts of anthropomorphism and zoomorphism (more generally, *animorphism*), the observed tendency of humans to identify or imagine life-like or human traits in non-human entities (Epley et al., 2008; Löffler et al., 2020) they observe, including abstract shapes (Heider & Simmel, 1944), inanimate objects (Burgess et al., 2018), animals (Epley et al., 2008), and robots (Złotowski et al., 2018). Some arguments posit that this tendency may be biologically grounded and instinctual, as even infants react to social robots as if they were

alive (Meltzoff et al., 2010), or may be based in psychological motivations, such as one's need for socialization and potentially inventing social actors (in this case, the social robot) to interact with and rationalize their environment (Epley et al., 2008). Thus this animorphization process may include both automatic and more deliberate (conscious) components (Złotowski et al., 2018), and may involve multiple dimensions of life-likeness (Złotowski et al., 2014).

Regardless of the underlying mechanism, evidence has mounted that people in practice do treat robots as life-like social entities (Złotowski et al., 2018), much more than with other interactive technologies such as personal computers (Nass & Moon, 2000; Seo et al., 2015), with a broad range of demonstrated effects including feeling socially obliged to assist robots (Sanoubari et al., 2019), engaging with rapport building behaviors with social robots (Seo et al., 2018), and many more. Thus, it follows that studying how people form expectations of other people they encounter can inform how we expect people to form expectations of social robots they interact with.

3.2 Fundamentals of Forming Expectations Between People

In this section, we review several major theories and models of human-human expectation formation, which we will in the following sections synthesize into a set of key points and a description of the cognitive process by which we anticipate that people will form expectations of robots. Specifically we will look at message passing models (primarily the *encoding/decoding model*; Hall et al., 1980), *expectancy violations theory* (Burgoon & Jones, 1976), *simulation theory* (Gordon, 1986), and *embodied interaction* (Dourish, 2001). We chose these

theories out of a larger, informal investigation into the topic of human-human expectations, selecting four which held particular salience for their application to human-robot interaction.

3.2.1 Message Passing

A predominant paradigm for analyzing inter-personal interaction (and sometimes with animals or robots) is message passing (Holthaus et al., 2023), where complex interaction is deconstructed into a serial set of discrete messages between the two (or more) interlocuters.[1] For example, the now-ubiquitous *encoding/decoding model* (Hall et al., 1980) breaks complex communication into a series of messages that are broadcasted by one party (e.g., spoken, facial expressions, gestures, etc., whether intentionally or not) and observed by a receiver (e.g., by listening or watching). Following, all messages go through multiple stages of abstraction before a receiver can interpret them: messages are encoded, sent (by the sender), transmitted through a medium (e.g., the physical world), received, and finally decoded (by the receiver), before one can make sense of them (illustrated in Figure 12).

Each phase provides an opportunity for the information to be altered, lost, or misconstrued (i.e., corrupted; Hall et al., 1980). The observer thus must rely on their particular imperfect

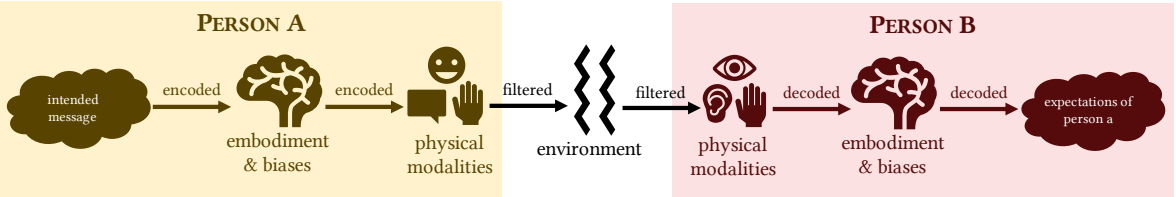

Figure 12: A message is passed from Person A to Person B only after being encoded by the sender's cognitive biases and physical form, filtered through the medium of the environment, and decoded by the recipient's own modalities and biases.

[1] An interlocutor is someone who takes part in an interaction.

decoding of messages, and not any necessarily true meaning or intent, to form expectations. For example, people may erroneously decode a scene and see faces in inanimate objects where none exist (*pareidolia*; Wodehouse et al., 2018), and develop inaccurate expectations of interaction. In this case, the receiver must resolve this expectation discrepancy using additional information.

This framing highlights several important points pertaining to constructing expectations of robots. First, we can dissect complex human-robot interactions into discrete units (e.g., a smile, a particular response, that a robot has hands) for targeted analysis regarding expectation formation. Second, we must assume that all information received is heavily filtered and modified from the transmission and receiving process; it is these imperfect messages, emitted by a robot, that shape expectations that people form.

3.2.2 Expectancy Violations Theory

More specific to our inquiry, *expectancy violations theory* (Burgoon & Jones, 1976) is a standard lens in communication studies for unpacking interaction between two people, that emphasizes how people hold and maintain expectations of an interlocutor as interaction unfolds or changes. Pre-existing or initial expectations (at interaction start) draw from the person's background and disposition, including social expectations and prior experience, whether in general, with the particular interlocutor, or with related entities. Following, as interaction unfolds new information often does not match existing expectations exactly, creating a *violation*, hence the name of the theory (Burgoon, 2015; Burgoon & Hale, 1988). Violations can be dramatic, such as an expected-to-be calm person becoming surprisingly

violent (Figure 13), but are typically more incremental, such as a person taking an unexpectedly informal and familiar tone given the professional relationship or situation, or even mundane and unremarkable, such as an unexpected switch in topic within a conversation.

Violations feed into iteratively evolving expectations: new information leads to expectations being revised rather than replaced. This means that expectations are relatively persistent, and for example may be based on pre-conceptions or individual prior experience (Burgoon & Hale, 1988). This highlights the importance of earlier expectations on interpreting violations. As an example, consider if a self-proclaimed topic expert (initial expectation) joins one's team, only to demonstrate moderate performance (violation); the updated expectation may be that the person has poor self-assessment or is dishonest. In contrast, if the person instead introduced themselves as a complete beginner (initial expectation) but then demonstrated the same still-unexpected moderate behavior (violation), one may instead lead to updated expectations of the person being modest or a fast learner. In this way,

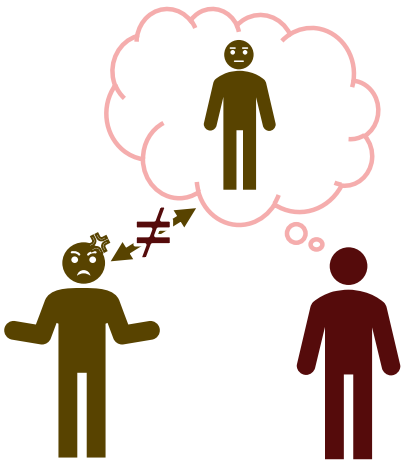

Figure 13: The observer (right) is startled when the subject (left), who they have previously known to be a calm, mild-mannered individual, suddenly behaves angrily and aggressively, violating their prior expectation.

expectation formation is reflexive: rather than being set according to most recent observations, expectations are the accumulation of incremental and continuous violations over a timespan.

For human-robot interaction then, given that we expect people to have less experience with robots, their predisposition towards technology and existing ideas (e.g., from media) may serve an outsized role in initial expectations. Further, these initial expectations are likely to be persistent, even as one interacts with a real robot, with expectations evolving incrementally over time as violations occur based on observations and interactions.

3.2.3 Simulation Theory

A complementary view on expectations is *simulation theory*, which postulates that people develop expectations of others through forms of mental state attribution (projecting mental states onto others; Shanton & Goldman, 2010), conducting internal cognitive simulations of how *they themselves* would behave given a similar situation (Gordon, 1986); mirror neurons, those that activate when observing an action as if one were doing the action, may be biological evidence of this (Gallese & Goldman, 1998; Shanton & Goldman, 2010). Simulation theory is in contrast to the idea that people more systematically apply logical rules and cognitive theories to develop their expectations of how others may behave (aptly called *theory theory*; Gordon, 1992). While both provide targeted lenses to consider, pragmatically we expect people to leverage a combination of simulations and internal theories to develop expectations to understand others' behavior.

These simulations are necessarily conducted from the observer's individual perspective, based on their own biases and leveraging known elements of the others' circumstances (Tamir & Mitchell, 2013) to achieve a plausible understanding (Epley et al., 2004). This explains known problems with expectations we may hold of others, including *naïve realism*, where people apply their own experience as an objective reality from which to understand others (Ross & Ward, 1996), and *realist bias* (Mitchell et al., 1996) or the *curse of knowledge* (Birch & Bloom, 2007), where a person assumes that the knowledge they hold is shared by others. These theories have been extended to non-human entities (e.g., animals, mechanical devices; Ames, 2004; Krueger, 2007; Meltzoff, 2007) and includes the proposal that anthropomorphism helps people fit observations into their existing knowledge to support simulation (Epley et al., 2007). Simulation theory supports our position that, due to animorphism, we may expect people to build expectations of robots as they do for other people. However, given the key differences between robots and people, we need to carefully consider what other inputs (e.g., robot design, previous knowledge of robots, etc.) may modulate the simulations of a robot's actions.

As an example of simulation theory, consider noticing someone litter even though they were standing near a salient garbage can. The observer may consider (simulate) what would lead *them* to litter next to a bin (Mitchell et al., 1996), only to conclude that the litterer is of poor moral character (Ross & Ward, 1996), based on their worldview against littering (Figure 14). However, suppose the observer knows the litterer personally and would expect better behaviour. This alternate perspective shapes the simulation, irrespective of the observation, and may instead lead them to acknowledge that the litterer did not notice or

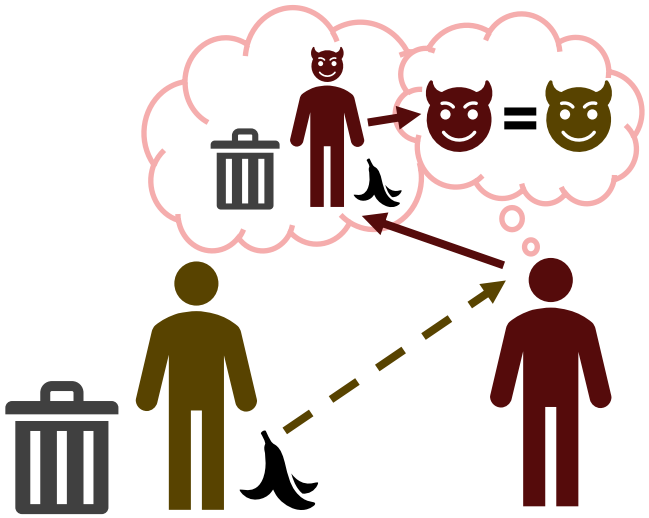

Figure 14: The observer (right) notices the subject (left) littering while standing next to a trash bin. The observer simulates themselves performing the same action, and concludes that they would only do so if they were malicious and immoral. They then extend this understanding of themselves in order to judge the subject as similarly immoral.

could not see the trash can, and update their expectation accordingly (Epley et al., 2004; Tamir & Mitchell, 2013). In either case, simulations are rooted in the perspectives of the observer (Gordon, 1992).

3.2.4 Embodied Interaction

Taking a step back, we highlight the importance of more broadly considering *embodied interaction* with respect to building expectations of robots (Dourish, 2001). From foundations in Heideggerian philosophy, concepts surrounding embodiment are central to communications studies (e.g., see Streeck et al., 2011), with *embodied interaction* (commonly discussed in human-computer interaction; Dourish, 2001) taking a phenomenological approach. Embodiment focuses on the role of a person's body and existence within the world (tangible, social, etc.) as foundations of cognition and interaction. All interactions between a person and an other (whether human, animal, or robot) must be mediated through one's

embodiment in the world, their *structural coupling* with their environment (Ziemke, 2003) (Figure 15). In other words, a person's experience (expectations, simulations, interpretations, etc.) cannot be decoupled from their body (size, shape, abilities, senses) and social reality (race, gender identity, nationality, background, etc.).

Embodiment thus provides a foundation for understanding all the theories presented, highlighting the critical role of one's own embodiment in message interpretation, expectation violations, and simulation theory. All interpretation and consideration is foundationally biased from an individual's own perspective, regardless of any external reality (e.g., about the robot's capabilities). Taking this to logical extremes, *symbolic interactionism* argues that people act according to an understanding of an object rather than the object as it truly is, embedded within the context in which the person and robot exist (Hoggenmueller et al.,

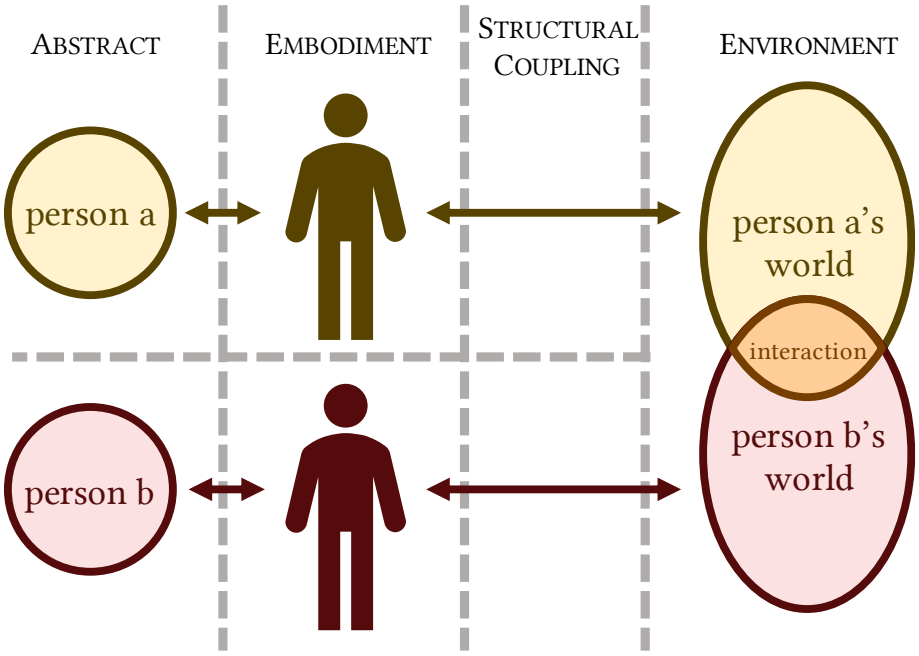

Figure 15: Embodied interaction between two people where each party is physically and socially embodied and structurally coupled to the world. Interaction between the two parties can only occur at the intersection between their embodiments.

2020). We can even consider society itself to be constructed from embodied interpretations formed through interactions between people (Carter & Fuller, 2015).

3.2.5 Summary

In this section we introduced prevailing theories of how people develop expectations of other people, through a process of passing and interpreting messages and using the encoded information to build and iterate upon expectations of others. This may include iteratively updating expectations (through violations) and be driven by cognitive simulations of how one would act (simulation theory), but all expectations will be developed from an individual's highly biased perspective. In the following sections, we synthesize these ideas into a cognitive process that explains how a person may form expectations of a robot they encounter.

3.3 Synthesis of Expectation Formation for Robots

We summarize the above discussion, synthesizing with specifics of human-robot interaction, into a set of key points for understanding how we may expect people to form, update and maintain their expectations of a robot over time. In the following subsection, we will further synthesize this discussion into an overarching Cognitive Process of Human-Robot Expectation Formation that can be used to analyze interaction and providing insight into the root of expectations that are formed, and thus the cause and potential solutions for expectation discrepancies encountered.

3.3.1 Individual Perspectives Dominate Expectations

First we highlight how embodied interaction means that people can only interact with machines from within a very narrow conceptual overlap between their personal complex physical and social contexts (Young et al., 2011), and, the robot's own presence within the world (Ziemke, 2003). This means that observations, messages, and violations are heavily translated using one's unique biases, world view, etc., and also that information the observer can receive *from* the robot is limited to a narrow overlap of embodiments (Figure 16); a person will develop expectation based on robot capabilities that they can both observe, and, make sense of within their embodiment. For example, it does not matter if a robot has cloud computing capabilities or the ability to recognize faces if the person cannot observe or understand this, even subtly (e.g., as in Thiessen et al., 2019). We cannot expect people to self-educate, necessarily ruminate on or try to untangle their observations, or consider what the robot can actually do more generally. Thus, our first key point in understanding expectation formation process is that

> **individual perspectives dominate expectations**

more than any underlying reality.

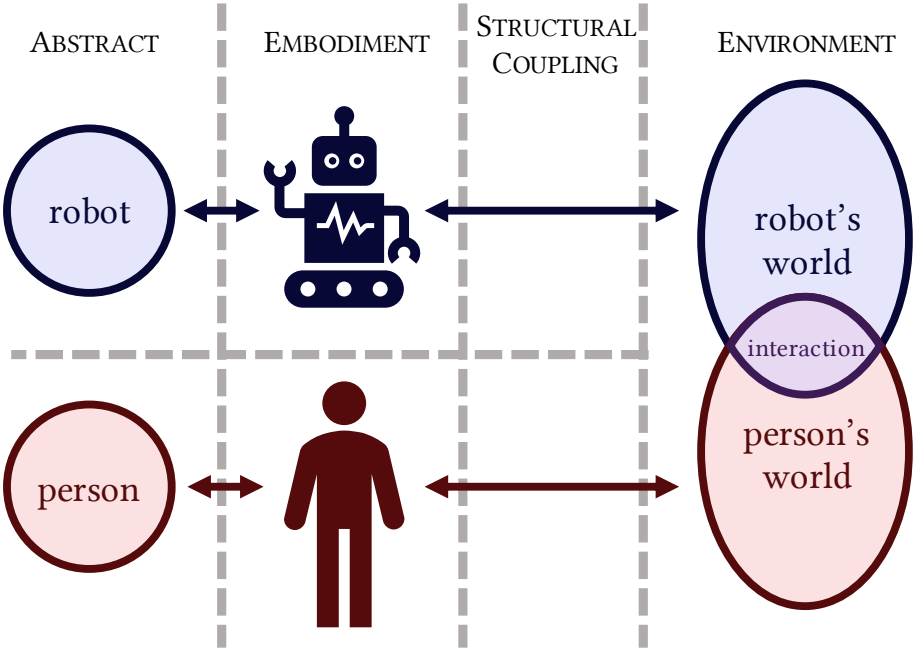

Figure 16: Embodied interaction between a person and a robot where each party is physically and socially embodied and structurally coupled to the world. Interaction between the two parties can only occur at the intersection between their embodiments.

3.3.2 Robot Designers Have Limited Direct Influence

Regardless of what a designer may intend for people to expect of their robot, the encoding/decoding model (Hall et al., 1980) highlights that any designed or intended signals (robot design, behaviours, etc.) must go through a complex transmission, translation, and interpretation process before they are interpreted and understood (Figure 17). After all, we expect people to react to a robot based more on their understanding than any reality of the

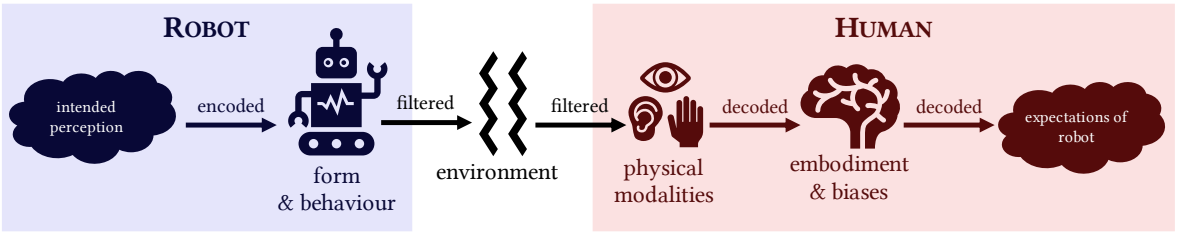

Figure 17: Any objective robot reality is translated and filtered, with many opportunities for alteration and error, and highly biased by the user, before it feeds into building a person's understanding and expectation of the robot.

robot's capabilities or intentions (Hoggenmueller et al., 2020). With robots, the prevalence of fantastical media depictions may lead to robots being seen predominantly as cultural concepts, with signals being interpreted in this light, more so than as technological objects (Hannibal, 2023; Richardson, 2015); we cannot expect clear distinctions between fact and fiction when people develop expectations of robots (Hannibal, 2023). For example, even if a robot designer tries to make a robot look like it cannot walk (e.g., by not having legs), a person may misinterpret and assume the robot has hidden wheels below it, based on expectations of robots being mobile. All of this emphasizes the fact that

> **robot designers, and the features, visual designs, behaviours, etc. that they create, have limited direct influence on expectations.**

Instead, designers need to accept that their creations may not be received as intended and consider their robot's designs and behaviors within the context of how people will interpret them.

3.3.3 People Make Sense of Robots in Terms of Themselves

Animorphism, embodied interaction, and simulation theory all suggest that people will understand social robots and build expectations as if the robot were alive; people have biological and social tendencies toward animorphism, understand other agents by simulating actions for themselves (Gallese & Goldman, 1998; Shanton & Goldman, 2010), and have mirror neurons that may activate when observing a robot (Gazzola et al., 2007; Hoenen et al., 2016; Oberman et al., 2007). There is ample evidence of this both for robots with human-like designs (Gazzola et al., 2007; Oberman et al., 2007), and in those with inanimate designs but

in sympathy-inducing situations (Hoenen et al., 2016). Regardless of objective facts about a robot's workings we expect observers to apply naïve realism (Ross & Ward, 1996) and project (their own) human-like personal circumstances, reasoning, and motivations, onto robots to make sense of their observations and generate expectations and understanding (Figure 18). Thus, overall we expect that

people make sense of robots in terms of their own likely behaviour,

or at least, as a similar social entity (e.g., another person).

3.3.4 Expectations Are Biased Toward Initial Impressions

As emphasized by both simulation theory and embodied interaction, the processing of any information or signals one receives is influenced by their background and predisposition

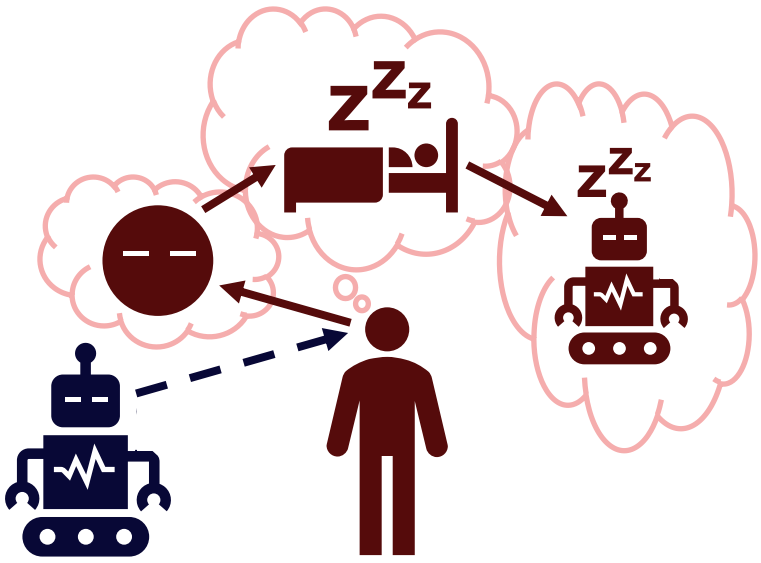

Figure 18: We expect people to make sense of observations using self-simulations based on what they see. Here, observing a robot with closed eyes, lowered head, and limp arms, a person simulates themselves and links to human sleep, concluding that the robot is in a sleep-like state.

toward that information. Notably, we expect people's predispositions to resist change, even in the face of new information (Tamir & Mitchell, 2013). As a consequence, new information does not directly lead to entirely new expectations, but rather, is processed within one's embodiment to update existing expectations. Any expectations we develop are generally resistant to change and iteratively reflexive, evolving with new information or violations instead of being replaced (Burgoon & Hale, 1988). It takes time and accumulated expectancy violations to shift existing expectations, even if they are quickly-adopted first impressions (Lemaignan et al., 2014). This is supported by evidence in HRI research, where first impressions can have a lasting effect (J. Xu & Howard, 2018) and impressions of robot capabilities evolve with repeated interactions (Paetzel et al., 2020; Figure 19). Therefore, in order to understand how an observer will process signals from a robot, we must understand what existing or prior understanding the person holds. This is exemplified by the *pratfall effect*,

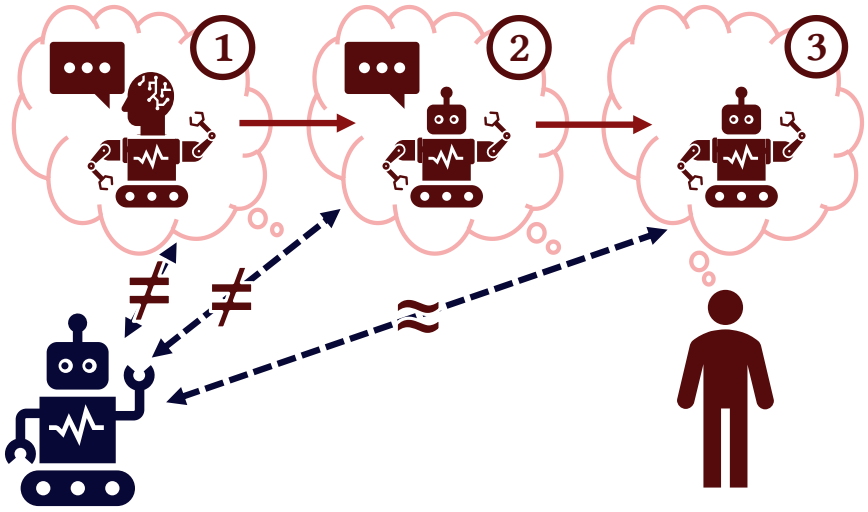

Figure 19: Expectations evolve during interaction, starting from a-priori beliefs; new information *modifies* existing expectations. For example, an observer (1) seeing a humanoid assumes the intelligent interaction ability, (2) after poor conversation behavior lowers expectations but still assumes it can talk, (3) after continued poor ability they no longer expect it can talk. Initial expectations thus change gradually, rather than simply being overwritten.

where a person may be seen as more likable when they make mistakes, *if* the person was previously seen as competent (Aronson et al., 1966). Conversely, a person who was previously seen as incompetent, upon making the same mistake, may now be seen as less likable; this has been observed in human-robot interactions (Mirnig et al., 2017). Thus

> **expectations are biased toward initial impressions and are updated (not replaced) by new information,**

and are therefore relatively persistent and resistant to change.

3.3.5 Summary

The fundamental property highlighted by all of these key points is the extensive conceptual distance between the reality of a robot (its capabilities) and any expectations that people form about the robot, with many steps of indirection, translation, and interpretation from an individual's perspective. These serve to inform our model of the cognitive process of human-robot expectation formation, which we present in the following section.

3.4 A Cognitive Process of Human-Robot Expectation Formation

We culminate our above discussion into a detailed process to describe and analyze how we expect a person to develop and maintain their expectations of a robot they encounter or interact with. This process is constructed as a composite of the four key points, connecting them together into a larger, more complete model of expectation formation. The process proceeds as follows:

A robot emits signals such as visual and behavioral design (which may not relate to actual capabilities) that the person observes. Simultaneously, there are additional external signals that provide exposition, e.g., any introduction to the robot or context (e.g., robot is in a factory). The observer receives these signals from within their physical and social embodiment, applying biases that shape their information interpretation and processing. Following, we expect the observer to simulate what observations of the robot would mean for them, promoting animorphic interpretation. All of this feeds into an evolving expectation of the robot, heavily influenced by earlier expectations and predisposition, to continuously update (likely persistent) expectations. These expectations feed back into shaping a person's long-term experiences, such that prior expectations, perhaps from previous interactions, start to influence new ones. This entire process is cyclic, with the inputs and resulting expectations continually evolving, as outlined in Figure 20.

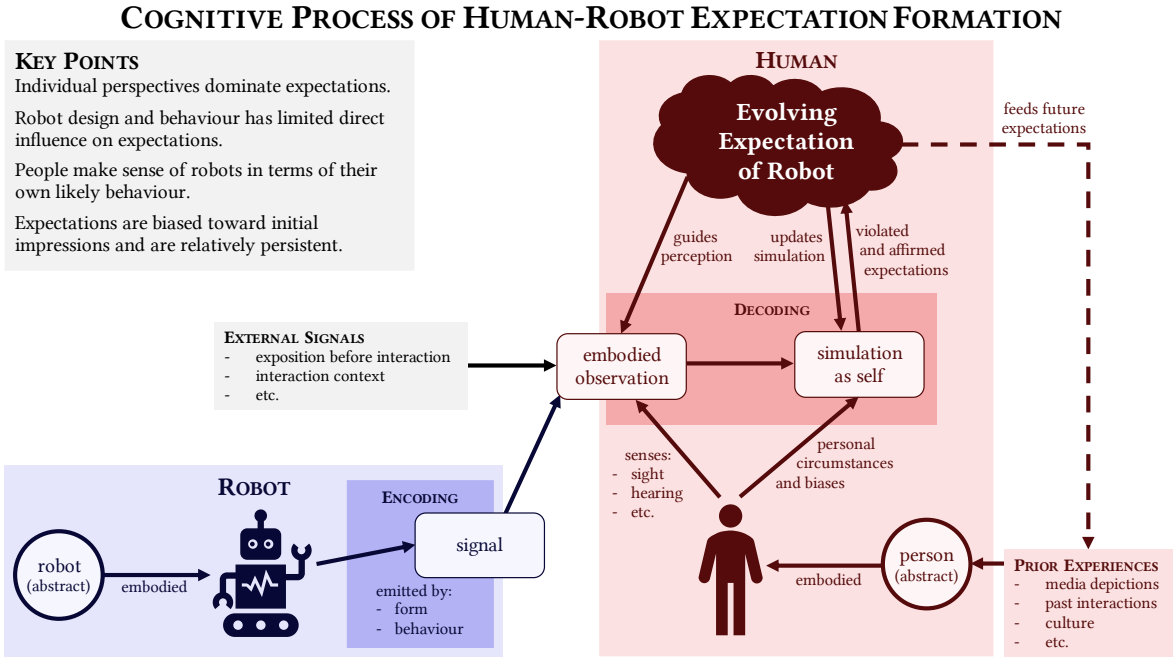

Figure 20: Our proposed Cognitive Process of Human-Robot Expectation Formation illustrating how people form and maintain expectations of robots they interact with.

Therefore, given the distance between robot capabilities and expectation formation, and how little control designers directly have over this, our cognitive process opens up this black-box in an attempt to decrease how much designers may perhaps "design and hope for the best" and instead provide a tangible series of steps, inputs, and cognitive elements. This offers robot designers an analysis tool that can be used to both help understand observed user expectations of past robots, as well as an exploratory guide to assist with predicting what expectations future users may form regarding the robot they are designing. Thus, while emphasizing their limited direct influence, we nonetheless offer designers a greater degree of control over and predictability over expectations of their robots, and through this support efforts to mitigate expectation discrepancy.

3.5 Chapter Summary

In this chapter, we aimed to explain how people form expectations of robots they encounter (RQ2). We began by reviewing prominent theories on expectation formation between people, which we used as a foundation for understanding people's expectations of robots (justified by the considerable evidence that people treat robot as if they were alive). We synthesized these theories from the perspective of human-robot interaction in order to develop a set of key points, which we used to model the overall cognitive process of human-robot expectation formation. This process model can be used by designers to help explain why users may be forming a particular expectation of their robot, and to highlight what factors may influence those expectations. In Chapter 5, we will demonstrate a technique for applying this process model in analyzing real robots, and in Chapter 6 we will reflect on that demonstration and critically evaluate the model's strengths and limitations.

Chapter 4 Classifying Expectations – Toward a Taxonomy

We have explained how expectations are formed, but up to this point, we have only spoken about the expectations themselves in broad terms. We have not yet examined what constitutes an expectation, or what 'expectation' even truly means. In this chapter, we will address RQ3: *What are the patterns in expectations that people form of robots, and can we distill them into a taxonomy?*

As noted in the introduction we use the term 'expectations' to refer generally to the beliefs a person holds about a robot's capabilities and behaviour. Our usage of the term contrasts with some other works in the field, which employ a stricter, more statistical definition regarding beliefs in the probabilities of future events (e.g. Rosén et al., 2022). We take this approach because it more accurately captures the discrepancies we target: prior literature on expectation discrepancy has focused on the presence of absence of qualities and capabilities rather than beliefs in the probability of a particular outcome (e.g., Kwon et al., 2016; Lohse, 2011; Schramm et al., 2020).

However, this broad (and thus not very specific) definition does not necessarily provide clarity regarding what one means when they say "expectations" of a robot. For example, if one expects that a robot has physical capabilities, does this mean they believe it can move around a room, or manipulate fine objects, or just change its shape? Similarly, if one expects that a robot has social capabilities, does this mean they believe that it can hold a conversation, or sense their emotional state, or has its own emotional state?

We develop a simple yet encompassing taxonomy to describe the various kinds of expectations that a person may form or hold of a robot or of their interaction with that robot. To achieve this, we take an inductive approach, surveying the field and available robots for both how the term expectations has been used, and, what expectations people may have of robots, analyzing these to develop key representative groupings. This results in an initial taxonomy, derived from the field, that provides vocabulary for more specifically discussing expectations of robots.

4.1 Process

The primary goal of developing our taxonomy was to create an initial vocabulary to explain the variety and depth of expectations that we were observing. As such, our process was less formal and did not include a full systematic review and instead was focused on reaching an initial taxonomy that provided a full coverage of the phenomenon that we observed or was noted in the literature.

We engaged with this process by consulting theoretical human-robot interaction literature on expectations, surveying pertinent studies of human-robot interaction (e.g., those dealing with expectations), and by collecting a breadth of representative robot platforms, prototypes, or behavior designs, designed for interaction with people. We further included prominent robots from science fiction. For this stage we conducted searches of academic sources using Google Scholar, and the ACM Digital Library. For our non-academic search we further used Google. In all cases we used keywords including ‘robot’, ‘expectation’,

'impression', and 'evaluation'[2]. This resulted in a corpus of images, videos, and behavior descriptions representing a range of robots and interactions.

We analyzed this corpus to extract expectations that people form or we suspect they may form, and to enumerate the robot or interaction characteristics that may contribute to the expectations (e.g., robot has eyes, or legs). This included directly drawing from literature, as well as informal brainstorming by our team (e.g., following analogs to our cognitive process) to uncover the range of potential factors and expectation outcomes. This resulted in a significant list of plausible expectations and robot design characteristics.

Finally, we thematically analyzed this collection using iterative, inductive processes, aiming to simultaneously cover works found while using as simple of a categorization scheme as possible, resulting in our taxonomy. Specifically, following an initial review of our expectations list, we constructed broad initial categories and began to organize them according to common patterns. We took inspiration from common qualitative analysis methods such as affinity diagramming (e.g. Harboe & Huang, 2015) to support our thematic analysis. As we classified the expectations, we iteratively adjusted these categories and reclassified expectations until we settled on a set of categories that described the collected examples as succinctly as we could find. We present this set in the following section.

[2] This was meant to capture users' and participants' evaluations of robots.

4.2 An Initial Expectations Taxonomy

Our process resulted in a two-dimensional taxonomy of expectations, which describes and categorizes the full range of expectations we uncovered through our exploration. One dimension is *expectation capability domains*, which includes only three nominal categories: physical, social, and computational expectations. The other dimension is *expectation abstraction*, which includes four ordinal categories from simple to more complex abstractions: rudimental, operational, purposive, and characteristic. To assist in introducing the dimensions of this framework, we will use the SoftBank Pepper (Aldebaran, n.d.) and Sony aibo (*Aibo*, n.d.) as example robots (Figure 21).

4.2.1 Domains of Expected Capability

The first dimension emerging from our thematic analysis classifies expectations into broad domains of capability. We identified three primary groupings of expectations of robot capabilities:

Figure 21: The SoftBank Pepper (Aldebaran, n.d.) and Sony aibo (*Aibo*, n.d.) used as examples throughout this section.

Physical Capabilities – People form expectations about how a robot may interact with and move within its physical environment. This can include expectations of the robot's movement abilities, such as expecting that Pepper can wave its arms (Figure 22), or that aibo can walk across the room. This also includes other outputs such as the ability to emit light or sound, as well as sensory capacities such as expecting that a robot can or cannot see, feel, touch, or receive radio waves.

Social Capabilities – In treating robots as social actors, people form expectations about a robot's abilities to communicate socially, as well as to integrate with and participate in society. For example, people may expect that Pepper can speak, hold a conversation, do social gestures (such as a wave or high five), or pay attention to a person (Figure 23). They may believe aibo can interpret facial expressions and infer emotional states, and possess its own internal emotional state as well. People may further expect that the robot can parse interpersonal relationships, understand the social dynamics in a group, or participate in social conventions such as yielding access to an elevator when socially appropriate (Aj. Moon et al., 2016).

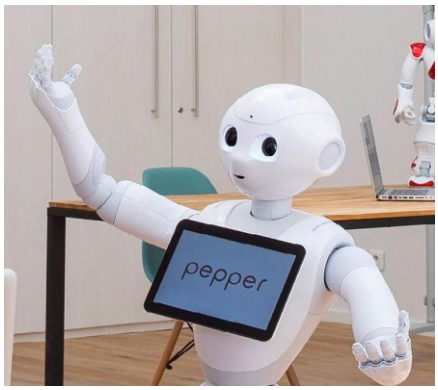

Figure 22: Pepper's (Aldebaran, n.d.) humanoid form can imply (correctly) that it can move its arms around to gesture, although its hands may imply more manual dexterity than it truly possesses.

Figure 23: Pepper's (SoftBank Robotics America, Inc., n.d.) face tracking behaviour may give the impression that the robot is paying attention to a person, regardless of whether it can really hear or understand anything being said to it.

Computational Capabilities – People form expectations about a robot's ability to think, in a computational sense. This generally encompasses a similar range of expectations to those people can form of a traditional computer. For example, they may expect that Pepper can perform mathematical or logical calculation, or that aibo can remember their face (Figure 24). Computational expectations can also include a robot's access to information sources (e.g., databases, encyclopedias, etc.), or whether it is capable of learning.

Figure 25 provides a visual summary of these domains. Note how the boundaries are blurred; this indicates how expectations can sometimes hit multiple categories or be highly linked. For example, the expectation that aibo can learn to perform dog tricks relates both to its physical abilities (to perform the movement) as well as its computational abilities (to learn and remember the tricks), while the belief that Pepper will shake a person's hand is both physical and social.

Figure 24: The intimacy aibo (*Aibo*, n.d.) displays toward users may encourage the impression that it recognizes their face and remembers them.

4.2.2 Levels of Expectation Abstraction

Orthogonal to the expectation capability domains discussed above, expectations of robot capabilities tended to range from purely mechanical capabilities (e.g., a motor can move; Cha et al., 2015) to more abstract, complex behaviors, and even robot intentions and personalities (e.g., Kuzminykh et al., 2020). We place these expectations on an ordinal dimension with four levels of abstraction, where at each level we could anticipate expectations on any of the three domains (physical, social, computational).

Rudimental Expectation – People form expectations of the basic mechanical capabilities of a robot such as the belief that Pepper can speak, and perform calculations, or that aibo's

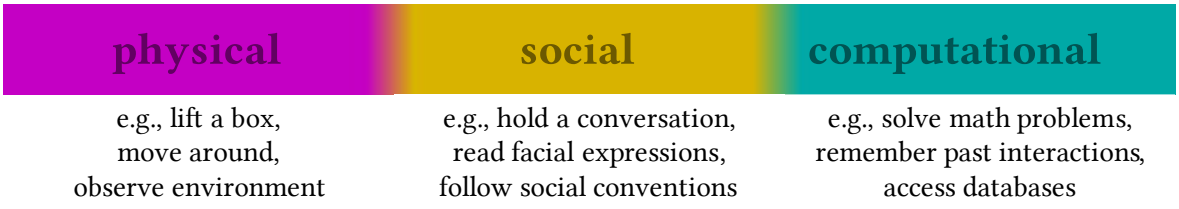

Figure 25: Examples of expectations falling into each of the three expectation capability domains. Note that the boundaries between domains are blurred, and it is possible for some expectations to lie across them.

legs have motors that are strong enough to move it. This is the expectancy of raw capability, not how a robot may be able to use it to perform operations.

Operational Expectation – People form expectations that a robot can use its rudimental capabilities to perform specific operations in its real-world environment. For example, a person may expect that Pepper can use its speech ability to engage in a conversation about a person's day, or that aibo's motors will enable it to climb over a box that is in its way.

Purposive Expectation – People will form expectations of a robot's goals and what actions it may take to meet those goals; this is in contrast to capabilities which they may believe a robot has but may not necessarily perform. For example, a person may expect that while Pepper *could* chat with them about their day, it is in a professional setting and will choose not to. On the other hand, if they expect that aibo wants to navigate across the room, they may expect it to climb over any obstacles it is capable of traversing. Whether based on an animorphic view of desires or plain expectations of an algorithm, this adds a layer of intentionality to otherwise mechanical robotic behaviors.

Characteristic Expectation – Just as with animals or other people, observers may attribute to a robot general traits or qualities, analogous to a personality. This includes general, high-level assessments of the robot's 'character' or personality, including mechanical impressions such as aibo being strong and capable, as well as human-flike impressions such as Pepper being friendly but professional.

Figure 26 provides a visual summary of these levels of abstraction. Note that unlike the blurred boundaries in Figure 25, the boundaries between abstraction levels are clearly defined, as we did not find in our corpus expectations that could not be cleanly organized into a singular category. We expect considerable interaction and interplay between the layers of abstraction, particularly adjacent ones. For example, if a person expects that a robot has eyes and can see (rudimental), this may lead them to assume that it can also recognize people (operational) and is trying to monitor them (purposive). Inversely, if a person holds a more abstract expectation, such as a robot being chatty (characteristic), this may infer expectations at lower levels, such as that the robot wants to talk to them (purposive), is able to hold a conversation (operational), and has the mechanisms to emit noise (rudimental). Further, there are negative cases where holding one expectation (e.g., that a robot is greedy) creates a negative expectation as a consequence (e.g., the robot will not give a cookie). We note that these trains of logic may not be supported by the reality of robot capability, resulting in an expectation discrepancy.

4.2.3 A Two-Dimensional Taxonomy of Expectations

Taken together, our capability domains and levels of abstraction form a two-dimensional taxonomy of what expectations people may develop for social robots. This framing enables

rudimental	operational	purposive	characteristic
e.g., has motors, can calculate, makes noise	e.g., lift a box, solve a math problem, hold a conversation	e.g., wants boxes, wants to avoid people, wants to comfort	e.g., friendly, greedy, chatty

Figure 26: Examples of expectations representing each of the four levels of expectation abstraction.

us to both categorize expectations and position how they relate to one another. At any point in the classification space, one could imagine an expectation that has both a capability domain (physical, social, computational) as well as a level of abstraction (mechanical, personal, etc.).

To visualize this space, we use a polar diagram, plotting the domains of expected capability as colours along the angular axis and the levels of expectation abstraction along the radial axis (Figure 27). A key interaction we found between the dimensions is that the domain becomes more difficult to classify at higher levels of abstraction. For example, rudimental expectations such as the ability to move, calculate, or speak are easily classified into the physical, computational, and social domains respectively. More complicated actions at the operational level, such as giving a hug, become somewhat harder to classify, blurring the line between the physical and social domains. At the most abstract layer (characteristic), classifying capability domain becomes particularly frustrated: while a characteristic like 'strong' is clearly physical, 'greedy' could be said to be both computational and social, and 'youthful' does not clearly align with any domain of capability[3]. This interaction is visualized by the starkly divided colour regions at the core of the diagram (the rudimental layer) gradually blurring together and eventually to white at the outer region (the characteristic layer).

[3] Indeed, expectations at the characteristic level can sometimes stray from the concept of a capability entirely.

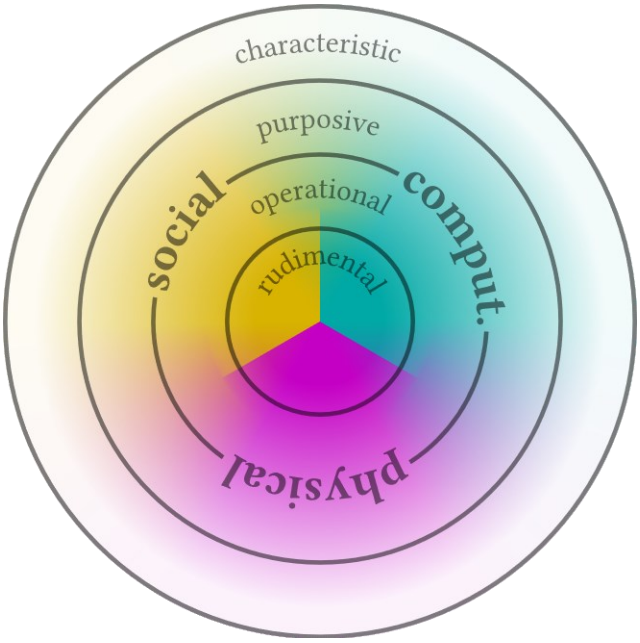

Figure 27: A two-dimensional taxonomy of expectations of robots, with capability domains on the angular dimension and levels of abstraction on the radial dimension. Note that the line between the capability domains blur as one moves further away from rudimental capabilities, as the higher-level expectations (e.g., that a robot is friendly) may involve multiple modalities. A user's set of expectations of a robot may be plotted on this diagram in order to visualize them and identify common areas of discrepancy, as demonstrated in Chapter 5.

4.3 Chapter Summary

In this chapter, we aimed to provide a vocabulary for describing and classifying expectations of robots (RQ3), which we developed in the form of a taxonomy of human-robot expectations. To develop this taxonomy, we collected a large corpus of expectations from literature in the field and conducted a thematic analysis to identify commonalities, ultimately organizing them into two dimensions. We then combined those dimensions into a two-dimensional classification space that can be used to visualize expectations and discrepancies. In Chapter 5, we will demonstrate a technique for applying this taxonomy in analyzing real robots, and in Chapter 6 we will reflect on that demonstration and critically evaluate the model's strengths and limitations.

Chapter 5 Demonstration with Analytical Techniques

In Chapters 3 and 4 we developed tools to support researchers and designers in engaging with the problem of expectation discrepancy, but there remains a gap between the theoretical tools and how to apply them in practice. Drawing from existing standard methodologies, we designed two potential analytics techniques to serve as examples for employing our framework: *systematic expectation dissection* and *cognitive expectation walkthroughs*. In this chapter, we present these methodologies and conduct case studies with each method to demonstrate their application. This demonstration further provides an opportunity to reflect more broadly on how our taxonomy and process model can be integrated into a design process. In doing so, we will address RQ4: *How can our improved knowledge of human-robot expectations be used by robot researchers and designers to examine and explain expectations of their robots?*

5.1 Systematic Expectation Dissection

We propose a novel methodology for analyzing observed or predicted user expectations of a robot. We call this new technique *systematic expectation dissection*: a guided exploratory process in which a designer organizes and identifies trends in user expectations by plotting them within our taxonomy of human-robot expectations. This technique produces a visual summary of user expectations and expectation discrepancies, and further prompts the designer to consider what types of expectations users are forming and how they may fit into a larger picture of how the robot is perceived. In this section, we will explain how to conduct an expectation dissection and demonstrate the process and outcome using case studies.

5.1.1 Visualizing Expectations

Our taxonomy can be used to visualize a user's expectations of a robot by identifying individual expectations and plotting them within the two-dimensional space. For example, the expectation that a robot can remember the user's face may be represented by a symbol in the operational layer, at a point between the social and computational capability domain (Figure 28).

To visualize expectation discrepancy, we need to further denote how expectations relate to a robot's true capabilities. Thus, we can mark an expectation plot point according to two additional properties, which we call *polarity* and *matching*. Polarity refers to whether an expectation is positive or negative: whether the user expects that the robot has a particular property or that it does not. Matching refers to whether an expectation aligns with a robot's underlying reality. For example, a mistaken belief that a robot can speak is a positive,

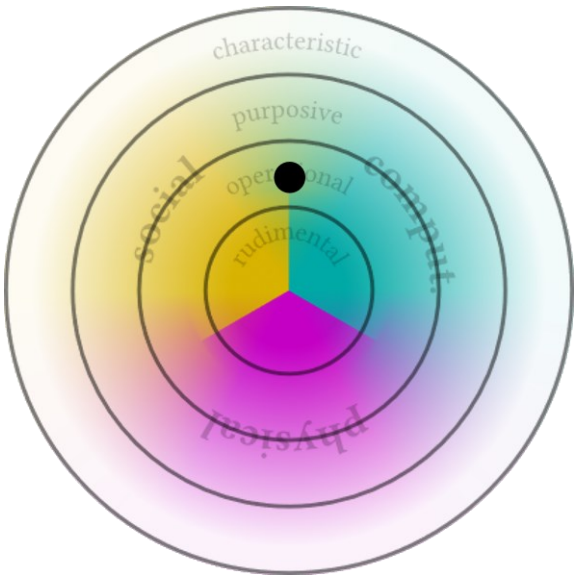

Figure 28: An expectation that a robot can remember a user's face is operational, and both computational and social, and so is plotted (with a dot) on the graph in the operational layer, along the blurred boundary between the computational and social regions.

mismatched expectation, while a correct belief that a robot cannot speak is a negative, matched expectation.

For illustrative simplicity, we treat each of these properties as binary. We mark the plot symbol for these properties, with the shape of the icon designating polarity (positive **+** vs. negative **–**) and the colour designating matching (**blue for matched** vs. **red for mismatched**). We note that in reality these properties are not binary, and in particular it may be difficult to classify a given expectation as simply matched or mismatched, especially with more abstract characteristic expectations; it may in practice be helpful to employ a more nuanced means of evaluating expectation matching. However, given our goal of facilitating a broad high-level overview of the expectations a person may hold of a robot, we believe that even a coarse-grained binary approach produces an effective visualization, and may allow for quicker evaluation of iterative designs.

For example, consider that a user believes a particular robot…

1. cannot speak (rudimental, social, positive, mismatched)
2. can remember a user's face (operational, social/computational, positive, matched)
3. is not well-informed (characteristic, computational, negative, mismatched)
4. does not want to move around (purposive, physical, negative, mismatched)

Using our taxonomy visualization, we can plot these expectations as in Figure 29. This produces a visual summary of user expectations. A designer or researcher can then look at the

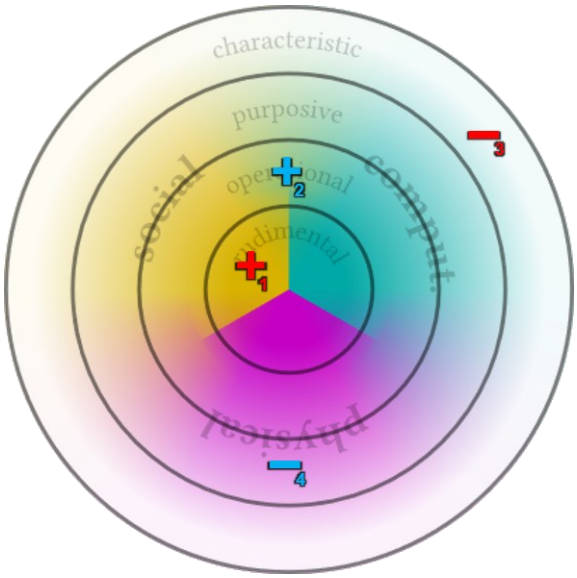

Figure 29: Example expectations visualized with our taxonomy of expectations using the plotting scheme explained in Section 5.1.1. The numbers on the symbols correspond to their position in the list at the end of the section.

graphic they have produced and see at a glance where user expectations are focused, and where there may be areas of strong expectation matching or discrepancy.

5.1.2 Procedure

To conduct a *systematic expectation dissection* using this taxonomy, one starts by assembling a list of expectations people may hold about their robot. How this list is compiled is outside the scope of this technique; the objective is simply to collect a wide range of plausible expectations of the robot. For example, they may compile this list by showing study participants a picture of the robot and simply asking them what they expect of it. There plenty of other possible methods, including analogizing from observations and study results of similar robots, or perhaps even by critically examining the design themselves and informally predicting what a person may expect when interacting with the robot.

While many methods may be employed, the overall goal of the exercise is to support a designer or researcher in engaging with the full range of potential expectations. The outcome of this process is a listing of expectations similar to the examples listed at the end of the previous section (Section 5.1.1).

Once a list of expectations is compiled, the method proceeds by plotting each of expectations onto the taxonomy space. This plotting procedure will require designers to think closely about each expectation, with consideration on how to classify them. While categorization of expectations within the taxonomy is somewhat subjective and will doubtlessly vary from designer to designer, the results in aggregate will nonetheless present a graphical summary of potential user expectations of the robot, as seen in Figure 29 above.

Designers can then examine this summary to identify patterns, in particular by searching for concentrated areas of matched or mismatched expectations, which may suggest key strengths and weaknesses of the robot's design with respect to user expectations.

5.1.3 Case Studies: Systematic Expectation Dissection

We demonstrate this approach with three real example robots as case studies: the SoftBank Pepper (Aldebaran, n.d.) , the Sony aibo (*Aibo*, n.d.), and SnuggleBot (Passler Bates & Young, 2020). To help demonstrate the technique, we informally generated example expectations that a hypothetical user may have of these robots (Table 1). Rather than a replacement for other experimental inquiry, this technique is intended as a form of heuristic exercise, with the goal of supporting researchers and designers in engaging with the full range of potential expectations which may emerge from their robot.

Our first example robot is the SoftBank Pepper (Aldebaran, n.d.). We took our hypothetical user expectations (Table 1) and plotted them onto our taxonomy space (Figure 30); the visual overview provides quick insight into common expectation patterns in the form of

No.	SoftBank Pepper	Sony aibo	SnuggleBot
1	can do addition	affectionate	can communicate with lights
2	can gesture	can bark	cannot do math
3	can give a hug	can do simple dog tricks	cannot have a conversation
4	can have a conversation	can jump	cannot move body
5	can move from place to place	can know if a person is in front of it	cannot move from place to place
6	can notice gestures	can learn	cannot understand speech
7	can speak	can remember my face	comforting
8	can speak French	can understand dog commands	cuddly
9	cannot compute an integral	can walk	does not have a camera
10	does not have specific knowledge	cannot speak English	does not have a microphone
11	empathetic	friendly	does not want to move
12	friendly	has camera	durable
13	has camera	has microphone	has buttons to press
14	has microphone	has speakers	has lights
15	intelligent	loyal	makes sounds
16	not well-informed	robust	not intelligent
17	not very strong	wants to approach people	soft
18	wants to answer questions	wants to move around the room	wants to comfort
19	wants to approach people	wants to seek attention	warm
20	wants to avoid collisions	young	
21	wants to help		
22	wants to invite interactions		
23	wants to shake hands		
24	won't bump into me		

Table 1: A set of hypothetical expectations for three different robots generated by the researchers. This list is not provided as empirical data about the robots, but rather as example data to be used to demonstrate how our taxonomy can visualize a user's expectations. The number in each row corresponds to the labeled plot symbols in the example visualizations (Figures 30-32).

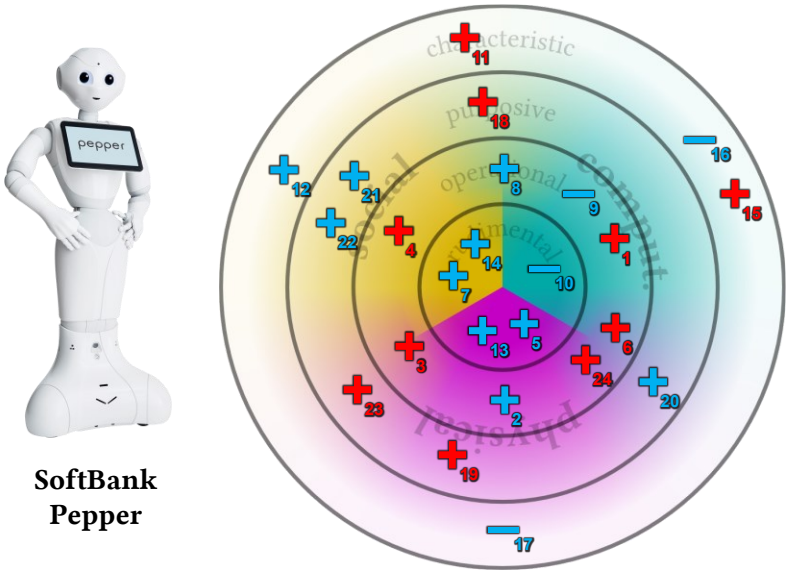

Figure 30: Example expectations (Table 1) of the SoftBank Pepper (Aldebaran, n.d.) visualized with our expectations taxonomy.

clusters of plot points, as well as conspicuously empty regions. As Pepper is a highly configurable robot, we consider a typical, largely 'default' configuration for the purpose of determining whether a particular expectation is matched or mismatched. One standout feature of Pepper's expectation visualization is that mismatched expectations are scattered fairly evenly across the domains and levels, with the notable exception that there were no mismatched rudimental expectations. While the user has a seemingly accurate understanding of Pepper's rudimental capabilities (e.g., they understand that it possesses a camera and that it has the ability to move around), they have mismatched expectations of how it will behave in practice (e.g., they mistakenly believe its ability to see means it will not bump into them as it moves about the area). This implies that Pepper encourages a wide range of expectation discrepancies, rather than being localized to any particular function or feature.

Our next example robot is the Sony aibo a robotic dog designed to fulfill the role of a pet in a user's home (*Aibo*, n.d.). We again plotted the expectations in Table 1 onto our taxonomy space (Figure 31). When comparing the expectation visualization for aibo to that of Pepper, it is immediately clear that the expectation discrepancies are more localized in nature. In particular, most of the mismatched expectations are abstract and either physical or social in nature. This includes assuming dog-like physical and social capacities that aibo does not really possess nor imitate (e.g., seeking out people, loyalty to one's owner).

Our final example robot is SnuggleBot (Figure 32), a stuffed narwhal with lights, mobile limbs, and sensors, which is designed to provide companionship to users (Passler Bates &

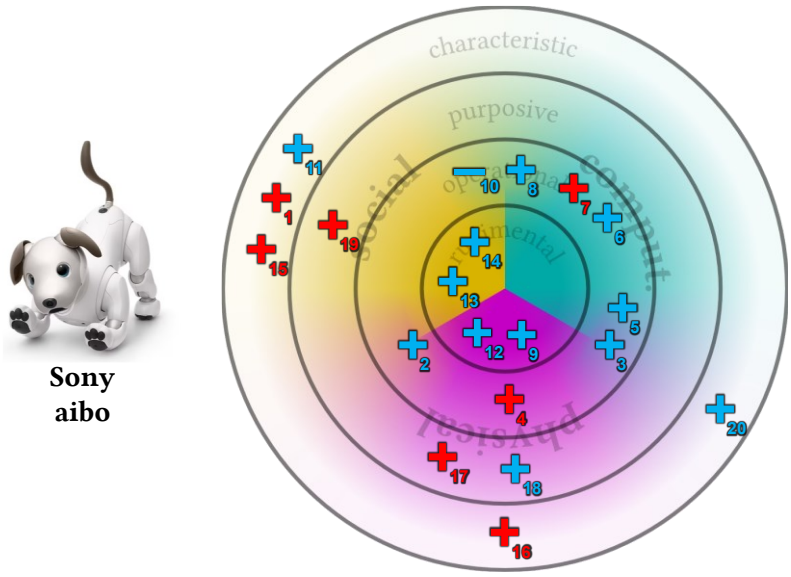

Figure 31: Example expectations (Table 1) of the Sony aibo (*Aibo*, n.d.) visualized with our expectations taxonomy.

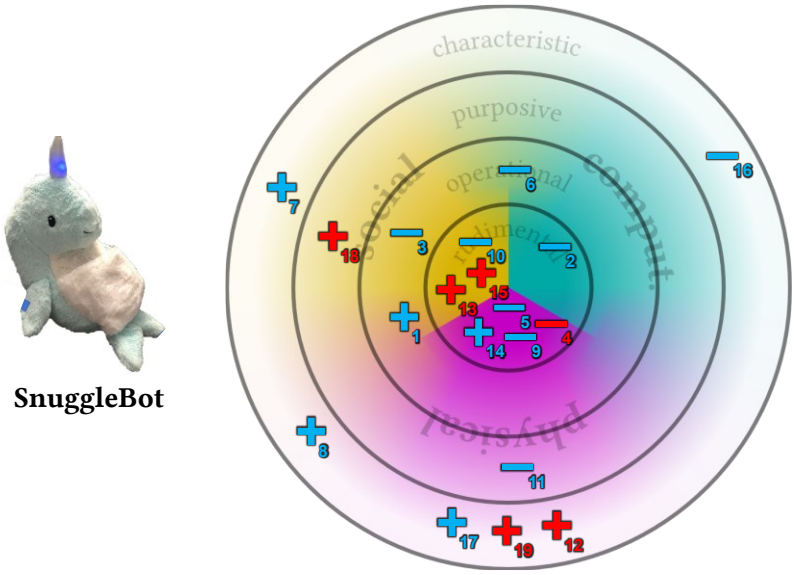

Figure 32: Example expectations (Table 1) of the SnuggleBot (Passler Bates & Young, 2020) visualized with our expectations taxonomy.

Young, 2020). One immediate difference with this visualization is that, compared to the other two robots, the user possessed many more negative expectations (expectations that the robot did not possess various capacities), perhaps because of the robot's simpler appearance resembling a stuffed animal. Further, many of the user's mismatched expectations are at the rudimental level, suggesting that the robot's appearance may be misaligned with its basic mechanical capabilities (e.g., the user does not expect that the limbs can move, but does expect that it will make sounds).

5.1.4 Summary

Systematic expectation dissection is a potential technique to support researchers and designers in comprehensively exploring the full range of potential expectations and expectation discrepancies with their robots. By classifying and plotting the expectations onto our taxonomy space, they produce a visual summary of what kinds of expectations users may

hold, and how those expectations compare, at a high level, to the robot's real abilities. Such a visualization can offer an initial guide toward necessary areas of focus for mitigating expectation discrepancy.

5.2 Cognitive Expectation Walkthroughs

Our expectation dissection allows designers and researchers to identify areas of expectation discrepancy, but does little to help explain or understand those discrepancies. This is where designers can employ our cognitive process for human-robot expectation formation, using an adaptation of standard Human-Computer Interaction analytical evaluation methods to conduct a *cognitive expectation walkthrough*: a scenario-based walkthrough of a specific human-robot interaction[4].

Continuing from one of our case studies in the previous section, we present an example of an expectation walkthrough with the SoftBank Pepper. We envision that after performing a systematic expectation dissection on their robot and uncovering major expectation discrepancies, a researcher or designer may next perform a cognitive expectation walkthrough to better explain those discrepancies and gain insights on how they may then alter the design to mitigate the issue.

5.2.1 Procedure

Our expectation formation process provides a cognitive framing for understanding how a person may develop and maintain expectations of a robot they encounter. Given a persona

[4] Although similar in terminology, we note that this is a distinct adaptation of the standard "cognitive walkthrough" methodology.

and a scenario (e.g., following standard human-computer interaction methodology; Rogers et al., 2023), a designer could conduct a cognitive expectations walkthrough using our expectation formation process to focus on elements of the interaction, context, and robot design (i.e., visual design, behavior, etc.) that impact expectations, how a person may interpret these to form and update their expectations, and how expectations evolve over time as the interaction unfolds.

Central to cognitive expectation walkthroughs is the development of a persona to represent a user, in addition to the interaction scenario and context, given the importance of an individual's background, biases, etc., in shaping expectations. We suggest unfamiliar readers to consult human-computer interaction texts for more details on persona development (e.g., such as Rogers et al., 2023). A cognitive expectation walkthrough also requires an established or proposed clear robot design and behavior; given the importance of robot features, it is challenging to do more generic robot-agnostic, walkthroughs. Thus, to conduct a walkthrough, there are three key components: the robot platform and behaviour we are examining, a potential user persona whose perspective we will adopt, and a scenario that gives context to and drives the interaction.

Following, we conduct the walkthrough by simulating, step by step, how the interaction may unfold. We analyze the encounter by consulting our model of the cognitive process for how we expect the person to maintain and build their expectations (Chapter 3). We systematically consider all the signals the person receives, including those from the robot's form and behaviour, as well as from the context of the interaction, and trace them through

the cognitive process in order to evaluate how this may impact their evolving expectations. This offers a grounded, thorough perspective to examine why the person may form certain expectations and not others.

5.2.2 Case Study: Scenario Parameters

Robot — For this demonstration we continue with the SoftBank Pepper (Aldebaran, n.d.) robot, as a widely used representative social-robot humanoid. It will be running industry-typical kiosk-style software that does basic conversation and information delivery.

Persona — Our fictional user is Sam, a mid-20s Canadian student who identifies as female, is generally friendly, and has an interest in novel technologies (is a self-described "nerd"). Sam has never interacted with a robot before, but has often seen them on the news and pays particular attention in media.

Scenario — Sam has just encountered Pepper as a retail assistant in a department store, and has approached Pepper for assistance in finding the shoe department. In this case, Pepper is located near the front of a store next to a sign saying "I can help!", and is programmed with a standard kiosk-style information application, using speech and hand gestures to deliver information; it receives input via a few pre-selected buttons on the tablet (Figure 21). There is a small sign next to the tablet instructing people to touch it to start.

5.2.3 Case Study: Cognitive Expectation Walkthrough

Here we present the results of a cognitive expectation walkthrough as performed by the author, using our key parameters from the previous section.

When Sam first notices the robot, Pepper is looking around the room and moving its arms casually. The form and behavior signal a modern-looking physical design with a humanoid form made of shiny white plastic and a tablet computer, with eyes (with cheery lights), ears, a mouth, and articulated arms with movable hands. The robot is making a soft whirring noise (a fan) and the joints emit mechanical noises when moving. Simultaneously, external signals influencing the interaction include Sam noticing the "I can help!" sign (exposition signal), and immediately recognizing the robot from the news (media depiction signal).

From an embodied observation point of view, Sam notices the visuals more than the audio given the noisy scenario. Sam applies her existing experience of seeing the robot on the news to interpret these signals, and combined with her existing expectations of robots (animorphic) she did not notice the tablet computer as an interaction modality. Her interest in technology amplified her interest and attention, helping her focus on the robot's attempts at gesturing and communication. Given these observations, Sam's mental simulation as if she were the robot results in expectations suggesting, that due to the combination of human-like facial features, humanoid form, and moving parts, the robot likely has a range of familiar, human-like social capabilities.

Sam approaches the robot and waves, saying hello. The robot does not respond. Observing this signal with her existing expectations, Sam is surprised. Simulating this reaction, Sam initially wonders if the robot is simply unfriendly, violating her expectations, but then realizes the robot maybe did not hear her. Sam still expects that the robot can hear and converse with her. Several seconds later, the robot looks at Sam, and its eyes blink. Sam notices

this, and still expecting the robot to converse, this signal feeds into Sam's simulation to indicate that the robot is now paying attention. Sam quickly says hello again, but while talking, the robot interrupts Sam to say "Hello! How can I help you?" in a loud voice. This startles Sam, and violates her assumption that the robot was paying attention. This again feeds into her simulation, initially indicating that the robot may be friendly but perhaps has poor social etiquette. This further violates Sam's expectation of conversation ability, and Sam reduces her expectation of the conversation ability. Sam responds by saying that she is doing well, but Pepper again ignores Sam. Sam is starting to feel frustrated at the rudeness, and this violation further reduces her expectations of behavioural conversation ability. Sam repeats herself, but is ignored again. Finally, Sam feels that the robot is not friendly and may be ignoring her. At this point Sam notices the instructions telling her to touch the screen to start (external exposition signal), which is a strong signal that updates Sam's simulation to suggest that, after all, the robot may not have conversation ability. Sam is disappointed by this expectation discrepancy and starts to wonder if the robot can hear, and begins to doubt other robot capabilities.

Sam touches the screen and a menu appears with a selection of store departments. Simultaneously Pepper cheerfully says "I am happy to help you!" while gesturing exuberantly. The social signals are highly salient, drawing Sam's attention away from the tablet. These behaviors again feed into Sam's simulation, and violates her expectations that the robot *cannot* converse. Sam ignores this, but finds it difficult to resist trying to talk to the robot again. This pattern continues as Sam navigates the menus, Pepper talks and gestures cheerfully, and Sam tries not to respond to the social gestures. Sam's friendly personality feeds

into her embodied observation of this behavior, and she starts to feel as if she is being rude to the robot. Sam finds the information she was looking for.

Sam touches a visible "I'm done" button on the kiosk to finish her session. Pepper cheerfully says "Thank you, come again!" Sam interprets this signal, and her updated simulation makes her wonder if her expectations are incorrect: perhaps Pepper can converse? Sam says, "Thanks Pepper, I'll come again!" and waits, but Pepper never responds. This signal pushes Sam to solidify her low expectations of the robot, and to feel that social robots can be quite rude and inconsiderate. This entire interaction feeds back into Sam's overall expectations about robots, and will shape her future interactions with them.

5.2.4 Summary

Cognitive expectation walkthroughs are potential tool for supporting designers and researchers in engaging at a deeper level with users' expectations of their robots. They can take the expectations identified and summarized through a systematic expectation dissection and, using our model of the cognitive process of human-robot expectations, trace how these expectations may be formed and understand what factors may be contributing to them. By encouraging a researcher or designer to systematically look at each step of the interaction, following through the elements of our process model and using the vocabulary of our framework, they can develop a deeper understanding of how an expectation evolves, which can offer direction for how the robot's design may be altered to mitigate unwanted expectations and expectation discrepancies.

5.3 Chapter Summary

In this chapter, we presented and demonstrated two novel techniques which show how the tools we developed can be applied to understand user expectations and discrepancies with real robots, using case studies as examples. In doing so, we showed that we have developed an analytical framework (Chapters 3, 4) that can be used to examine and explain expectations that people form of robots (RQ4). Thus, we have demonstrated through these examples the potential that our framework has for supporting understanding of what people expect of their robots, and of the factors that contribute to those expectations, enabling designers and researchers to engage with and mitigate challenges related to expectation discrepancies. In Chapter 6, we will critically reflect on these techniques, as well as their respective tools, in order to evaluate our analytical framework, identifying strengths and limitations of these approaches and highlighting opportunities for future work.

Chapter 6 Critical Reflection

In the preceding chapters, we presented the framework we developed to support designers and researchers in examining people's expectations of robots. In this chapter, we perform a critical analysis of our framework to consider its merits and limitations, both with respect to its theoretical foundations and its practical use. Our demonstration applying our framework to case studies in Chapter 5 serves as a focal point for reflection on the framework more broadly.

We reflect critically on the components of our framework, based on our experiences developing it and applying it to case studies. Overall, while we made progress toward our research questions, in this chapter we take a critical view toward to the limitations of the work, including both the boundaries of the theoretical perspectives that inform it as well as challenges to applying it in practice. We organize these limitations according to salient themes that emerged from this analysis.

These limitations further inform key recommendations we make in Chapter 7, both for how our framework may be applied currently, as well as to highlight opportunities for future work to complement and build upon our tools.

6.1 Scope and Granularity

Our two-dimensional taxonomy offers a scheme for quickly classifying a wide range of expectations. In this section we discuss its wide scope and consequent loss of granularity, as well as how that scope relates to prior systems for measuring user expectations.

6.1.1 Broad Coverage of Expectations

In developing our taxonomy, we attempted to classify the full range of expectations found within the corpus we developed (see Section 4.1). We did not find any expectations in our survey or case studies which could not be placed within the bounds of the scheme.

While all expectation instances fit within the framework, some were more difficult to classify in that they fit into multiple regions of the space. The broad definitions of the categories within the taxonomy, designed as such to accommodate the immense variety in user expectations, resulted in a substantial degree of overlap between them. This was most prominent with the domains of capability, where expectations could frequently fall into multiple categories, particularly at higher levels of abstraction. For example, the expectation 'wants to shake hands' involves both a physical and social component. We represented this in the taxonomy visualization with increasingly blurred boundaries between the domains at the higher levels (see Section 4.2.3 for more details).

To a lesser extent, this overlap was also found across some levels of abstraction. For example, in one of our case studies, the expectations 'can speak' is treated as a rudimental expectation while 'can speak French' is treated as operational. The reasoning for this is that speech itself is regarded as a basic mechanical function whereas speaking a particular language is regarded as a specific operation. This classification is not obvious; there is a degree of subjectivity in how a designer using of the framework interprets the different levels. While we did not find this ambiguity to be a major obstacle to the overall goal of concisely summarizing expectations, it nonetheless raises the cognitive load of the task.

6.1.2 Taxonomic Space Collapses Differences

The broad scope of our taxonomy came at the cost of granularity in how it represents expectations. When using our taxonomy to compare and relate expectations, we note that it adopts a very generalized view of expectations, which results in a loss of nuance. For example, consider two expectations for a pair of robots: that Robot A can walk across a room and that Robot B can see an item on a shelf. Both of these expectations fall squarely into the same location in our taxonomy (physical, operational), yet they are very distinct and not directly related to one another. Thus, when conducting a systematic expectation dissection, the visualizations for the two robots will appear similar in that area of the space, despite not necessarily having anything in common.

The dimensions of our taxonomy have the potential to collapse important distinctions between expectations. When employing our framework, it is thus important to remember that it is but one generalized lens, and that despite shared patterns, expectations are ultimately particular to every individual robot and interaction.

6.1.3 Taxonomy Compared to Prior Frameworks

When comparing our taxonomy to prior works which map out dimensions of impressions of robots, we find a noteworthy distinction. The RoSAS scale (Carpinella et al., 2017) identified the dimensions of warmth, competence, and discomfort, the Godspeed scale (Bartneck et al., 2009) identified anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, and several others present similar sets of dimensions (Dupree & Fiske, 2017; Kuzminykh et al., 2020; Nomura et al., 2008). Positioning these works within our

taxonomy highlights that they all broadly deal with our outermost layer of expectation abstraction, characteristic expectations.

Our taxonomy thus expands on these existing works by integrating the layer of abstraction to help relate these characteristic expectations to underlying mechanical and behavioral elements. At the same time, these other works offer a more granular approach to measuring user impressions of robots at the characteristic level.

Overall, our framework serves to help designers in identifying patterns across the broad range of expectations and explaining why they form. This contrasts with and complements the above scales, which serve to measure particular impressions and expectations, as well as with the previously-discussed (see Section 2.3.2) Social Robot Expectation Gap Evaluation Framework (Rosén et al., 2022), which offers a mechanism to detect expectation discrepancy. Thus, we can imagine using all of these tools in concert in order to detect and explain user expectations of robots and understand how they relate to one another.

6.2 Foundations in Theory

Our framework was developed primarily through the consultation of prior literature. Our model of the cognitive process of human-robot expectation formation was synthesized from theories of expectations between humans, combined with existing literature on expectations of robots, while our taxonomy was mainly developed through thematic analysis on expectations of robots found in prior works. Neither of these components have been tested empirically with real data, which means they rest on some assumptions.

6.2.1 Cognitive Process Model

We developed our model of the cognitive process of human-robot expectation formation by collecting and synthesizing prominent theories on how expectations are formed and maintained between people, together with prior human-robot expectations literature.

This foundation in human-human interaction rests on the assumption that people will respond to robots in ways that resemble how they respond to people. As we discussed in Section 3.1, there is considerable evidence to support this idea: *animorphism*, the tendency to attribute life-like traits to non-living entities, is well-established in human-robot interaction. At the same time, this assumption somewhat contradicts the conclusion we came to in Section 2.1, where we identified the properties of a social robot that make interaction with them both similar and ultimately distinct from interaction with people.

While we attempted to address this contradiction through our combination of human-human and human-robot expectations literature, the model is nonetheless based on this assumption of animorphism. In our cognitive expectation walkthrough case study, for example, we found ourselves wondering whether the fictional user in the scenario would really treat the robot in such human terms, including regarding the robot as "friendly but [having] poor social etiquette". While such human-like attributions are well-attested in literature (see Section 2.1 for examples), it remains a leap to simply assume them when analyzing an interaction.

6.2.2 Expectations Taxonomy

Our taxonomy was developed through thematic analysis on our corpus of surveyed and hypothetical interactions. This provided a broad coverage of expectations in the field, but is nonetheless a limited qualitative analysis. One avenue for future work may be to conduct experiments measuring the expectations people form of robots to test whether they align with the dimensions and categories of our taxonomy, and ultimately develop a quantitative scale to measure user expectations along these dimensions.

6.3 Reliance on Designer Expertise

Our demonstration of preliminary application techniques on case studies (Chapter 5) highlighted how our framework can be applied effectively in order to examine user expectations of robots, but it also highlighted obstacles in the process. A key weakness of our framework is that applying it to evaluate real robots heavily relies on the expertise of the evaluators. Without hard data to draw from, our proposed application techniques rely on evaluators to use their own knowledge and judgement. In this section, we consider how our framework leans upon a designer or researcher's expertise, but also how it supports them in applying that expertise.

6.3.1 Systematic Expectation Dissection

In generating the informal example expectations to demonstrate systematic expectation dissection with our taxonomy (see Section 5.1), we note that it required substantial effort in order to generate expectations that approximately spanned the full space of the taxonomy. While there were plausible expectations spanning the full space, it was far easier to

think of certain types of expectations (especially rudimental) over others. This may suggest that, when applying the framework to analyze real user expectations, similar difficulties may be faced when interviewing a participant to collect their expectations. Thus, work remains to determine how best to interview people in order to extract the full range of their expectations.

Despite this difficulty, it is noteworthy that our taxonomy offers a guide on what types of expectations to specifically inquire about. This does not replace the need for a designer's expertise in order to describe and classify expectations, but it demonstrates the utility of our taxonomy in supporting them in that process.

6.3.2 Cognitive Expectation Walkthroughs

Our framework provides leads and perspectives for designers and researchers to explore and examine expectations of their robots. Our cognitive process model offers key insights for explaining why a user may be forming a particular expectation. For example, it emphasizes that expectations are evolving expectations and are resistant to change, which forces us to consider how previous expectations, both from earlier in the interaction and from a person's past experience, shape ongoing interactions. It also places heavy emphasis on mental simulation, which keeps us grounded in how a person engages in sensemaking of their observations, in contrast to our own technologist viewpoint.

Conducting cognitive expectation walkthroughs using our process model further emphasized these factors, making it easier to explain and how and when violations happen. Note how in our walkthrough case study (see Section 5.2), Sam did not readily form new

expectations given new information, but we would expect them to cling to earlier expectations. This illustrates the potential of our process model, using walkthroughs to explore and probe problem cases ("Why are people expecting this from my robot?"), or in the design process for a new robot ("What may people expect, and why?", "What if I change this?").

While, in the above ways, cognitive expectation walkthroughs using our process model support an evaluator in examining a design and explaining potential user expectations, the evaluator must nonetheless rely on their own expertise and judgement. That a user may form their expectations through mental simulation does not on its own tell an evaluator why a particular expectation is formed. It remains incumbent on the evaluator to explain these things themselves, but our framework serves as a guide to support them in doing so.

6.3.3 Framework is Not Predictive

It is essential to note that our framework does not predict expectations given some design. It does not directly explain what features of a robot contribute to which particular expectations, but instead serves as a probing tool to support designers in finding those causal connections. This serves as a foundation for designers to assert greater control over expectations and mitigate expectation discrepancies.

While an applicable predictive tool would be a great asset for designers and researchers in engineering desired expectations in users, we do not consider it to be within the scope of this work. Instead, our framework may offer a basis for the development of such a tool, providing necessary theoretical grounding and vocabulary.

6.4 User as Passive Observer

Our model of the cognitive process of human-robot expectation formation provides a high-level summary of how signals from a robot and interaction context combine with a person's internal state to result in expectations. In constructing this model, we necessarily emphasized certain aspects and deemphasized others. In particular, the process model treats the user as a passive participant, receiving signals and processing them in a predictable manner. While this perspective is grounded in the human-human expectation literature we reviewed, it is nonetheless just one perspective.

Furthermore, our model places a heavy emphasis on mental simulations: all inputs to expectations must go through the simulation step. As we discussed in Chapter 3, there is considerable debate regarding the degree to which mental simulations determine expectations, as opposed to a more rational, rules-based approach. For example, mental simulation may be an imperfect model for how people understand observations that are truly alien to their internal experience, but which they are nonetheless familiar with (e.g., computational capabilities).

The limitations of these theoretical perspectives are exemplified in our cognitive expectation walkthrough case study, where the focus of our analysis is on the internal cognitive processing of the fictional user. Our fictional user struggles through the interaction, confused by the robot's actions, but seldom takes action to rectify her confusion, instead relying only on iterative mental simulations and adjusting her conduct accordingly. In reality, we might expect that the user more actively seeks knowledge about the situation, prodding

the robot to explore its abilities or asking another person for help. Our cognitive process model provided a helpful guide for understanding the interaction, but it is not a definitive perspective and must be combined with others for a more holistic understanding.

6.5 Framework Generalizability

An additional limitation of cognitive expectation walkthroughs as an analytical technique is that they consider expectations somewhat specific to the given scenario and interaction, which makes it difficult to compare expectations (and thus expectation discrepancy) more generally between scenarios, interaction instances, and robots. This represents a tension between the importance of context, background and embodiment when understanding an interaction (see Section 3.2.4), and the ideal of developing a systematic understanding of expectations of robots across interactions.

We aim to address this tension by complementing the finer-grained understanding of expectations offered by a cognitive expectation walkthrough with the higher-level summary visualization produced through systematic expectation dissection, using our taxonomy of expectations to describe the general forms and types that we may expect people to form. Our taxonomy provides a unified vocabulary and broad scope for designers to describe and compare a person's expectations of a robot. Using case studies on three real robots, we demonstrated the power of this taxonomy to utilize a spatial presentation to highlight notable tendencies in user expectations and expectation discrepancies and compare them across the robots. It is important to note however, that such direct comparisons must be made with care for the nuances between similarly classified expectations (Section 6.1.2).

6.6 Chapter Summary

In this chapter we conducted a critical reflection on our framework in order to evaluate its utility to designers and researchers and to identify its limitations. We examined its scope and granularity, its foundations in theory rather than real data, its reliance on designer expertise for proper application, its treatment of the user as a passive observer in the interaction, and the challenges when using it to generalize about expectations across different interactions and robots. While we find our framework effective in addressing our research questions, it is critical to understand the limits of its application and perspectives. In the following chapter, we will use these limitations as the basis to develop recommendations for the successful application of our framework to analyze real robots, and to identify opportunities for future work to expand our knowledge of human-robot expectations.

Chapter 7 Conclusions

In this research, we developed an analytical framework to examine and describe people's expectations of robots. We demonstrated our framework using preliminary application techniques and reflected on it critically to identify its utility and limitations. In this chapter, we summarize our work and its implications. We begin by outlining our contributions and reflecting on how they relate to our research questions. We then review the limitations highlighted through our critical reflection. Considering these limitations, we provide some final recommendations for applying our framework in practice, and identify opportunities for future work. We conclude with a brief summary of our work and its role in addressing the problem of human-robot expectation discrepancy.

7.1 Contributions

In this work, we engaged with the problem of unpacking concepts surrounding people's expectations of robots they encounter and interact with. We then approached the problem from two major perspectives: explaining *how* people form expectations of robots through our model of the cognitive process of human-robot expectation formation, and enumerating *what types* of expectations they form through our two-dimensional taxonomy of expectations, together constituting an overall framework. We proposed preliminary techniques for how to employ these tools, and applied them to case studies as a demonstration of our framework. In this section, we reflect on our original research questions and consider how this work addresses them.

RQ1: How is the research community engaging with the concept of human-robot expectations?

Through our exploration of prior literature on the subject, we found that while substantial efforts have been made, there is a need for consistency in perspectives and vocabulary. This process highlighted areas of opportunity which we targeted with our framework.

RQ2: What is the process by which people form expectations of robots they encounter?

We conducted a background analysis and synthesis of prominent theories on expectation formation between people, resulting in a novel model of the cognitive process by which people form expectations of robots. This provides a breakdown of many of the important elements and inputs into this process. Grounded in a long history of human psychology and sociology, our process model offers a perspective rooted in a reality of how people think and engage in the world.

RQ3: What are the patterns in expectations that people form of robots, and can we distill them into a taxonomy?

We collected a large corpus of expectations and used thematic analysis to identify patterns, resulting in a two-dimensional taxonomy that classifies human-robot expectations according to domain of capability and level of abstraction. This taxonomy provides a concise means of describing and comparing human-robot expectations.

RQ4: How can our improved knowledge of human-robot expectations be used by robot researchers and designers to examine and explain expectations of their robots?

We developed two techniques: systematic expectation dissections and cognitive expectation walkthroughs, and applied them to case studies to demonstrate how our framework may be applied to understanding user expectations of real robots. This process of developing and testing these techniques further supported our evaluation, highlighting key advantages and challenges in applying our framework which we expand upon later in this chapter.

Altogether, we addressed each of the research questions we proposed at the beginning of this work, resulting in a framework that has advanced our understanding of human expectations of robots. In the following section, we discuss the key limitations of this work.

7.2 Limitations

Though we have made progress toward our research questions, our critical analysis of our framework in Chapter 6 identified several key limitations that must be considered. We summarize them here:

1. The broad scope of our taxonomy necessitated a coarse-grained representation of expectations that can obscure important nuances and distinctions.
2. Our framework is grounded predominantly in theoretical literature and relies on assumptions with limited empirical evidence.
3. Employing our framework relies heavily expertise, offering probing tools and guides but ultimately relying on the evaluator to make judgements.
4. Our model of the cognitive process of expectation formation does not sufficiently account for a user's active role in developing their understanding of a robot.

5. Information gained through cognitive expectation walkthroughs can be difficult to generalize beyond a particular interaction or robot.

These limitations do not prevent a designer or researcher from applying our framework in practice. Indeed, we found in our critical reflection considerable utility for them to employ our framework as a supportive probing tool. Rather, they must understand these limitations in order to employ the framework effectively. In the following sections, we draw from these limitations some recommendations for successful employment of our framework, and highlight opportunities for future work to fill in the gaps.

7.3 Recommendations

In light of the limitations of our framework, we have identified some key recommendations to guide designers and researchers in applying our framework to real robots in concert with other tools for understanding human-robot expectations.

7.3.1 Use as a Probing Tool

Our framework can be used as a probing tool to guide a designer or researcher in exploring a user's expectations of a robot, but cannot by itself provide direct, definitive answers on a how a user will respond to a particular design feature. Such answers may be approximated through an evaluator's expertise, but may require experimental methods in order to determine precisely.

7.3.2 Complement with Other Expectation Tools

Our framework can be used to complement other tools which detect and measure user expectations of robots. A designer or researcher may use such tools (discussed in Section 6.1.3) to identify expectations that users may hold, and then employ our framework to analyze and explain those expectations and understand how they relate to one another.

7.3.3 Remember Key Perspective Limitations

It must be remembered when employing our framework that it is limited by certain theoretical perspectives. The two most prominent of these are that our taxonomy offers limited granularity and can obscure nuanced differences between expectations, and that our model of expectation formation treats the user as largely passive in the process. These limited theoretical lenses, if not properly considered, may distort one's analysis of expectations.

7.4 Future Works

We note three major opportunities for future work to build upon and extend our framework. Future work in these areas may enable the development of practical, quantitative tools that designers and researchers could use to determine what features of their robot are leading to expectation discrepancies in users. Such tools would further empower them with control over user expectations, allowing them to mitigate discrepancy and achieve smoother human-robot interactions.

7.4.1 Amending Process Model with an Active, Rationalizing User

In Section 6.4 we emphasized that our model of the cognitive process of human-robot expectation formation is limited by its treatment of the user as a passive observer in the

process. This presents an opportunity to amend this model with a greater accounting for their active role in seeking out and understanding new information. This may include incorporating elements of *theory theory*, the rival to *simulation theory*, acknowledging the role that both may play in expectation development (more details in Section 3.2.3).

7.4.2 Empirical Validation of Taxonomy

In Section 6.2.2 we noted that our taxonomy of expectations was developed through thematic analysis on observed and hypothesized interactions and has not been experimentally tested. Moving forward, we believe it will be important for ongoing research to move beyond the theoretical and into more experimental work. This may take the form of experiments to measure the expectations people form of robots and determine whether the dimensions and categories of our taxonomy match the patterns of people's real expectations in practice, as well as whether our taxonomy space indeed covers all important cases. Such work may culminate in the development of a quantitative scale for measuring user expectations along the dimensions of our taxonomy.

7.4.3 Standardized Expectation Interview Methodology

In Section 6.3.1 we discussed the difficulties we encountered when generating hypothetical expectation data for our systematic expectation dissection case study (Section 5.1). This may imply that when interviewing a participant about their expectations, they may require coaching to elicit certain types of expectations that they not immediately think of. Thus, there is potential for a formal, standardized methodology for expectation interviews in order to extract a thorough listing of expectations from a participant. More research is

required to understand what types of expectations participants hold but may not readily express, and what questions may encourage them to express them.

7.5 Conclusion

The field continues to struggle with creating robot designs that lead users to expect unrealistically advanced, perhaps human-like capabilities in robots that are well beyond its capabilities, leading to disappointed and confused users, and perhaps failed robots. While designers continue to explore robots that are more transparent regarding their abilities, the field still lacks the knowledge necessary to support robot creators in making informed choices to influence users' expectations of their robots.

In this paper, we addressed four key research questions. We reviewed how human-robot expectations and expectation discrepancy are being discussed in research today, and identified a need for a consistent, systematic framework (RQ1). We presented a novel cognitive process we developed for how users form and maintain expectations of social robots (RQ2). We further developed and presented a taxonomy for describing and classifying expectations users may form according to key patterns (RQ3). Finally, we demonstrated preliminary analytical techniques for applying our framework to real robots, which illustrated the analytical power of our tools as probes that we envision designers can use in their own research and robot design processes as a basis to better anticipate and influence the expectations that their robots establish in users (RQ4).

As the field continues to improve our understanding of how to create robots that garner appropriate expectations, our work serves as an important step in providing concrete tools

to engage these problems. Ultimately, by enabling designers to more precisely influence user expectations, they may design robots that can better imply their true capabilities, mitigating expectation discrepancy and leading to more successful human-robot interaction.

References

Aibo. (n.d.). Retrieved 15 August 2022, from https://us.aibo.com/

Aldebaran. (n.d.). *Pepper the humanoid and programmable robot |Aldebaran*. Retrieved 29 September 2023, from https://www.aldebaran.com/en/pepper

Ames, D. R. (2004). Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. *Journal of Personality and Social Psychology*, *87*(3), 340–353. https://doi.org/10.1037/0022-3514.87.3.340

Aronson, E., Willerman, B., & Floyd, J. (1966). The effect of a pratfall on increasing interpersonal attractiveness. *Psychonomic Science*, *4*(6), 227–228. https://doi.org/10.3758/BF03342263

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *International Journal of Social Robotics*, *3*(1), 41–52. https://doi.org/10.1007/s12369-010-0082-7

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, 591–594. https://doi.org/10.1109/roman.2004.1374827

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Berzuk, J. M., & Young, J. E. (2022). More Than Words: A Framework for Describing Human-Robot Dialog Designs. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 393–401. https://doi.org/10.1109/HRI53351.2022.9889423

Birch, S. A. J., & Bloom, P. (2007). The Curse of Knowledge in Reasoning About False Beliefs. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

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

Brown, W. J. (2015). Examining Four Processes of Audience Involvement With Media Personae: Transportation, Parasocial Interaction, Identification, and Worship. *Communication Theory*, *25*(3), 259–283. https://doi.org/10.1111/comt.12053

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., & Tscheligi, M. (2013). The Good, The Bad, The Weird: Audience Evaluation of a "Real" Robot in Relation to Science Fiction and Mass Media. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social Robotics* (pp. 301–310). Springer International Publishing. https://doi.org/10.1007/978-3-319-02675-6_30

Bryant, D., Borenstein, J., & Howard, A. (2020). Why Should We Gender? The Effect of Robot Gendering and Occupational Stereotypes on Human Trust and Perceived Competency. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 13–21. https://doi.org/10.1145/3319502.3374778

Burgess, A. M., Graves, L. M., & Frost, R. O. (2018). My possessions need me: Anthropomorphism and hoarding. *Scandinavian Journal of Psychology*, *59*(3), 340–348. https://doi.org/10.1111/sjop.12441

Burgoon, J. K. (2015). Expectancy Violations Theory. In *The International Encyclopedia of Interpersonal Communication* (pp. 1–9). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118540190.wbeic102

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*, *55*(1), 58–79. https://doi.org/10.1080/03637758809376158

Burgoon, J. K., & Jones, S. B. (1976). Toward a Theory of Personal Space Expectations and Their Violations. *Human Communication Research*, *2*(2), 131–146. https://doi.org/10.1111/j.1468-2958.1976.tb00706.x

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 254–262. https://doi.org/10.1145/2909824.3020208

Carter, M. J., & Fuller, C. (2015). Symbolic interactionism. *Sociopedia*. https://doi.org/10.1177/205684601561

Cha, E., Dragan, A. D., & Srinivasa, S. S. (2015). Perceived robot capability. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 541–548. https://doi.org/10.1109/ROMAN.2015.7333656

Collins, E. C., Prescott, T. J., Mitchinson, B., & Conran, S. (2015). MIRO: A versatile biomimetic edutainment robot. *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology*, 1–4. https://doi.org/10.1145/2832932.2832978

Cross, E. S., & Ramsey, R. (2021). Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions. *Trends in Cognitive Sciences*, *25*(3), 200–212. https://doi.org/10.1016/j.tics.2020.11.009

de Graaf, M. M. A. (2016). An Ethical Evaluation of Human–Robot Relationships. *International Journal of Social Robotics*, *8*(4), 589–598. https://doi.org/10.1007/s12369-016-0368-5

de Graaf, M. M. A., Ben Allouch, S., & van Dijk, J. A. G. M. (2015). What Makes Robots Social?: A User's Perspective on Characteristics for Social Human-Robot Interaction. In A. Tapus, E. André, J.-C. Martin, F. Ferland, & M. Ammi (Eds.), *Social Robotics* (pp. 184–193). Springer International Publishing. https://doi.org/10.1007/978-3-319-25554-5_19

Dennler, N., Ruan, C., Hadiwijoyo, J., Chen, B., Nikolaidis, S., & Matarić, M. (2023). Design Metaphors for Understanding User Expectations of Socially Interactive Robot Embodiments. *ACM Transactions on Human-Robot Interaction*, *12*(2), 21:1-21:41. https://doi.org/10.1145/3550489

Dourish, P. (2001). *Where the Action is: The Foundations of Embodied Interaction*. MIT Press.

Dula, E., Rosero, A., & Phillips, E. (2023). Identifying Dark Patterns in Social Robot Behavior. *2023 Systems and Information Engineering Design Symposium (SIEDS)*, 7–12. https://doi.org/10.1109/SIEDS58326.2023.10137912

Dupree, C. H., & Fiske, S. T. (2017). Universal Dimensions of Social Signals: Warmth and Competence. In A. Vinciarelli, J. K. Burgoon, M. Pantic, & N. Magnenat-Thalmann (Eds.), *Social Signal Processing* (pp. 23–33). Cambridge University Press. https://doi.org/10.1017/9781316676202.003

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective Taking as Egocentric Anchoring and Adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When We Need A Human: Motivational Determinants of Anthropomorphism. *Social Cognition*, *26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*, 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Eyssel, F., Hegel, F., Horstmann, G., & Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. *19th International Symposium in Robot and Human Interactive Communication*, 646–651. https://doi.org/10.1109/ROMAN.2010.5598687

Fortunati, L., Manganelli, A. M., Höflich, J., & Ferrin, G. (2023). Exploring the Perceptions of Cognitive and Affective Capabilities of Four, Real, Physical Robots with a

Decreasing Degree of Morphological Human Likeness. *International Journal of Social Robotics*, *15*(3), 547–561. https://doi.org/10.1007/s12369-021-00827-0

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*(4), 1674–1684. https://doi.org/10.1016/j.neuroimage.2007.02.003

Glas, D. F., Kanda, T., & Ishiguro, H. (2016). Human-robot interaction design using interaction composer: Eight years of lessons learned. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 303–310. https://doi.org/10.1109/HRI.2016.7451766

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.*, 55–60. https://doi.org/10.1109/ROMAN.2003.1251796

Gompei, T., & Umemuro, H. (2015). A robot's slip of the tongue: Effect of speech error on the familiarity of a humanoid robot. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 331–336. https://doi.org/10.1109/ROMAN.2015.7333630

Gordon, R. M. (1986). Folk Psychology as Simulation. *Mind & Language*, *1*(2), 158–171. https://doi.org/10.1111/j.1468-0017.1986.tb00324.x

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

Hall, S., Hobson, D., Lowe, A., & Willis, P. (1980). Encoding/decoding. In *Culture, Media, Language* (pp. 117–127). Taylor & Francis Group. http://ebookcentral.proquest.com/lib/umanitoba/detail.action?docID=179321

Hamacher, A., Bianchi-Berthouze, N., Pipe, A. G., & Eder, K. (2016). Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 493–500. https://doi.org/10.1109/ROMAN.2016.7745163

Hannibal, G. (2023). *The Trust-Vulnerability Relation—A Theory-driven and Multidisciplinary Approach to the Study of Interpersonal Trust in Human-Robot Interaction* [Thesis, Technische Universität Wien]. https://doi.org/10.34726/hss.2023.108560

Harboe, G., & Huang, E. M. (2015). Real-World Affinity Diagramming Practices: Bridging the Paper-Digital Gap. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 95–104. https://doi.org/10.1145/2702123.2702561

Haring, K., Matsumoto, Y., & Watanabe, K. (2013). How Do People Perceive and Trust a Lifelike Robot? *Proceedings of the World Congress on Engineering and Computer Science (WCECS)*, *1*, 23–25.

Harris, J., & Sharlin, E. (2011). Exploring the affect of abstract motion in social human-robot interaction. *2011 RO-MAN*, 441–448. https://doi.org/10.1109/ROMAN.2011.6005254

Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., & Sagerer, G. (2009). Understanding Social Robots. *2009 Second International Conferences on Advances in Computer-Human Interactions*, 169–174. https://doi.org/10.1109/ACHI.2009.51

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259. https://doi.org/10.2307/1416950

Hoenen, M., Lübke, K. T., & Pause, B. M. (2016). Non-anthropomorphic robots as social entities on a neurophysiological level. *Computers in Human Behavior*, *57*, 182–186. https://doi.org/10.1016/j.chb.2015.12.034

Hoggenmueller, M., Chen, J., & Hespanhol, L. (2020). Emotional expressions of non-humanoid urban robots: The role of contextual aspects on interpretations. *Proceedings*

of the 9TH ACM International Symposium on Pervasive Displays, 87–95. https://doi.org/10.1145/3393712.3395341

Holthaus, P., Schulz, T., Lakatos, G., & Soma, R. (2023). Communicative Robot Signals: Presenting a New Typology for Human-Robot Interaction. *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 132–141. https://doi.org/10.1145/3568162.3578631

Hwang, K., & Zhang, Q. (2018). Influence of parasocial relationship between digital celebrities and their followers on followers' purchase and electronic word-of-mouth intentions, and persuasion knowledge. *Computers in Human Behavior*, *87*, 155–173. https://doi.org/10.1016/j.chb.2018.05.029

Jung, Y., & Lee, K. M. (2004). Effects of Physical Embodiment on Social Presence of Social Robots. *Proceedings of the 7th Annual International Workshop on Presence*, 80–87.

Kahn, P. H., Freier, N. G., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R. L., & Kane, S. K. (2008). Design patterns for sociality in human-robot interaction. *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 97–104. https://doi.org/10.1145/1349822.1349836

Kaminski, M. E., Rueben, M., Smart, W. D., & Grimm, C. M. (2016). Averting Robot Eyes. *Maryland Law Review*, *76*, 983.

Komatsu, T., Kurosawa, R., & 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), 109–116. https://doi.org/10.1007/s12369-011-0122-y

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology*, *18*(1), 1–35. https://doi.org/10.1080/10463280701284645

Kuzminykh, A., Sun, J., Govindaraju, N., Avery, J., & Lank, E. (2020). Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376665

Kwon, M., Huang, S. H., & Dragan, A. D. (2018). Expressing Robot Incapability. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 87–95. https://doi.org/10.1145/3171221.3171276

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. https://doi.org/10.1109/HRI.2016.7451807

Lemaignan, S., Fink, J., Dillenbourg, P., & Braboszcz, C. (2014). The Cognitive Correlates of Anthropomorphism. *HRI 2014 Workshop: HRI: A Bridge between Robotics and Neuroscience*. http://infoscience.epfl.ch/record/196441

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001

Löffler, D., Dörrenbächer, J., & Hassenzahl, M. (2020). The Uncanny Valley Effect in Zoomorphic Robots: The U-Shaped Relation Between Animal Likeness and Likeability. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 261–270. https://doi.org/10.1145/3319502.3374788

Lohse, M. (2011). Bridging the gap between users' expectations and system evaluations. *2011 RO-MAN*, 485–490. https://doi.org/10.1109/ROMAN.2011.6005252

Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, *10*(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, *23*(8), 966–972. https://doi.org/10.1016/j.neunet.2010.09.005

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI*, *4*. https://www.frontiersin.org/articles/10.3389/frobt.2017.00021

Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition*, *59*(1), 1–21. https://doi.org/10.1016/0010-0277(95)00683-4

Moon, Aj., Calisgan, E., Bassani, C., Ferreira, F., Operto, F., & Veruggio, G. (2016). The Open Roboethics initiative and the elevator-riding robot. In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), *Robot Law* (pp. 131–162). Edward Elgar Publishing. https://doi.org/10.4337/9781783476732.00014

Moon, Y. (2000). Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *Journal of Consumer Research*, *26*(4), 323–339. https://doi.org/10.1086/209566

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Natarajan, M., & Gombolay, M. (2020). Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 33–42. https://doi.org/10.1145/3319502.3374839

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of Human Behavior in Human–Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots. *IEEE Transactions on Robotics*, *24*(2), 442–451. IEEE Transactions on Robotics. https://doi.org/10.1109/TRO.2007.914004

Noor, N., Rao Hill, S., & Troshani, I. (2021). Artificial Intelligence Service Agents: Role of Parasocial Relationship. *Journal of Computer Information Systems*, 1–15. https://doi.org/10.1080/08874417.2021.1962213

Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing*, *70*(13), 2194–2203. https://doi.org/10.1016/j.neucom.2006.02.024

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In *Social psychology: Handbook of basic principles* (pp. 211–238). The Guilford Press.

Paepcke, S., & Takayama, L. (2010). Judging a bot by its cover: An experiment on expectation setting for personal robots. *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, 45–52.

Paetzel, M., Perugia, G., & Castellano, G. (2020). The Persistence of First Impressions: The Effect of Repeated Interactions on the Perception of a Social Robot. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 73–82. https://doi.org/10.1145/3319502.3374786

Passler Bates, D., & Young, J. E. (2020). SnuggleBot: A Novel Cuddly Companion Robot Design. *Proceedings of the 8th International Conference on Human-Agent Interaction*, 260–262. https://doi.org/10.1145/3406499.3418772

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is Human-like?: Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database. *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 105–113.

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016). Errare humanum est: Erroneous robots in human-robot interaction. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 501–506. https://doi.org/10.1109/ROMAN.2016.7745164

Reig, S., Forlizzi, J., & Steinfeld, A. (2019). Leveraging Robot Embodiment to Facilitate Trust and Smoothness. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 742–744. https://doi.org/10.1109/HRI.2019.8673226

Richardson, K. (2015). *An Anthropology of Robots and AI: Annihilation Anxiety and Machines*. Routledge. https://doi.org/10.4324/9781315736426

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. *Proceedings of the 4th ACM/IEEE*

International Conference on Human Robot Interaction, 245–246. https://doi.org/10.1145/1514095.1514158

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 101–108. https://doi.org/10.1109/HRI.2016.7451740

Robinson, F. A., Velonaki, M., & Bown, O. (2021). Smooth Operator: Tuning Robot Perception Through Artificial Movement Sound. *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 53–62. https://doi.org/10.1145/3434073.3444658

Rogers, Y., Sharp, H., & Preece, J. (2023). *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons.

Rosén, J., Lindblom, J., & Billing, E. (2022). The Social Robot Expectation Gap Evaluation Framework. In M. Kurosu (Ed.), *Human-Computer Interaction. Technological Innovation* (pp. 590–610). Springer International Publishing. https://doi.org/10.1007/978-3-031-05409-9_43

Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, *36*, 422–439. https://doi.org/10.1016/j.chb.2014.03.066

Ross, L., & Ward, A. (1996). Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. In *Values and Knowledge*. Psychology Press.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics*, *5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot

Cooperation and Trust. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 141–148. https://doi.org/10.1145/2696454.2696497

Sandoval, E. B., Mubin, O., & Obaid, M. (2014). Human Robot Interaction and Fiction: A Contradiction. In M. Beetz, B. Johnston, & M.-A. Williams (Eds.), *Social Robotics* (pp. 54–63). Springer International Publishing. https://doi.org/10.1007/978-3-319-11973-1_6

Sanoubari, E., Seo, S. H., Garcha, D., Young, J. E., & Loureiro-Rodriguez, V. (2019). Good Robot Design or Machiavellian? An In-the-Wild Robot Leveraging Minimal Knowledge of Passersby's Culture. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 382–391. https://doi.org/10.1109/HRI.2019.8673326

Schaefer, K. E., Sanders, T. L., Yordon, R. E., Billings, D. R., & Hancock, P. A. (2012). Classification of Robot Form: Factors Predicting Perceived Trustworthiness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1548–1552. https://doi.org/10.1177/1071181312561308

Schramm, L. T., Dufault, D., & Young, J. E. (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*, 439–441. https://doi.org/10.1145/3371382.3378280

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (2015). Poor Thing! Would You Feel Sorry for a Simulated Robot? A comparison of empathy toward a physical and a simulated robot. *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 125–132.

Seo, S. H., Griffin, K., Young, J. E., Bunt, A., Prentice, S., & Loureiro-Rodríguez, V. (2018). Investigating People's Rapport Building and Hindering Behaviors When Working with a Collaborative Robot. *International Journal of Social Robotics*, *10*(1), 147–161. https://doi.org/10.1007/s12369-017-0441-8

Shanton, K., & Goldman, A. (2010). Simulation theory. *WIREs Cognitive Science*, *1*(4), 527–538. https://doi.org/10.1002/wcs.33

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology*, *23*(3), 309–316. https://doi.org/10.1007/s10676-020-09573-9

Simion, F., Di Giorgio, E., Leo, I., & Bardi, L. (2011). The processing of social stimuli in early infancy. In *Progress in Brain Research* (Vol. 189, pp. 173–193). Elsevier. https://doi.org/10.1016/B978-0-444-53884-0.00024-5

Stanton, C. J., & Stevens, C. J. (2017). Don't Stare at Me: The Impact of a Humanoid Robot's Gaze upon Trust During a Cooperative Human–Robot Visual Task. *International Journal of Social Robotics*, *9*(5), 745–753. https://doi.org/10.1007/s12369-017-0422-y

Streeck, J., Goodwin, C., & LeBaron, C. (Eds.). (2011). *Embodied Interaction: Language and Body in the Material World*. Cambridge University Press.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*(1), 151–162. https://doi.org/10.1037/a0028232

Thiessen, R., Rea, D. J., Garcha, D. S., Cheng, C., & Young, J. E. (2019). Infrasound for HRI: A Robot Using Low-Frequency Vibrations to Impact How People Perceive its Actions. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 11–18. https://doi.org/10.1109/HRI.2019.8673172

van Maris, A., Lehmann, H., Natale, L., & Grzyb, B. (2017). The Influence of a Robot's Embodiment on Trust: A Longitudinal Study. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 313–314. https://doi.org/10.1145/3029798.3038435

Wodehouse, A., Brisco, R., Broussard, E., & Duffy, A. (2018). Pareidolia: Characterising facial anthropomorphism and its implications for product design. *Journal of Design Research*, *16*(2), 83–98. https://doi.org/10.1504/JDR.2018.092792

Xu, J., & Howard, A. (2018). The Impact of First Impressions on Human- Robot Trust During Problem-Solving Scenarios. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 435–441. https://doi.org/10.1109/ROMAN.2018.8525669

Xu, K., Chen, M., & You, L. (2023). The Hitchhiker's Guide to a Credible and Socially Present Robot: Two Meta-Analyses of the Power of Social Cues in Human–Robot Interaction. *International Journal of Social Robotics*, *15*(2), 269–295. https://doi.org/10.1007/s12369-022-00961-3

Yanco, H. A., & Drury, J. (2004). Classifying human-robot interaction: An updated taxonomy. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, *3*, 2841–2846. https://doi.org/10.1109/ICSMC.2004.1400763

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*, *1*(1), 95–108. https://doi.org/10.1007/s12369-008-0006-y

Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., & Grinter, R. E. (2011). Evaluating Human-Robot Interaction: Focusing on the Holistic Interaction Experience. *International Journal of Social Robotics*, *3*(1), 53–67. https://doi.org/10.1007/s12369-010-0081-8

Zafar, A. U., Qiu, J., & Shahzad, M. (2020). Do digital celebrities' relationships and social climate matter? Impulse buying in f-commerce. *Internet Research*, *30*(6), 1731–1762. https://doi-org.uml.idm.oclc.org/10.1108/INTR-04-2019-0142

Ziemke, T. (2003). What's that thing called embodiment? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25).

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

Złotowski, J., Strasser, E., & Bartneck, C. (2014). Dimensions of Anthropomorphism: From Humanness to Humanlikeness. *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 66–73.

Złotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C., & Ishiguro, H. (2018). Model of Dual Anthropomorphism: The Relationship Between the Media Equation Effect and Implicit Anthropomorphism. *International Journal of Social Robotics*, *10*(5), 701–714. https://doi.org/10.1007/s12369-018-0476-5