Expecting the Expected: An Analytical Framework to Examine People's Expectations of Robots

James M. Berzuk*, Ceilidh Simon, and James E. Young

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

berzukj@myumanitoba.ca * (ORCiD: 0000-0003-0242-4993)
ceilidh.simon@gmail.com
young@cs.umanitoba.ca (ORCiD: 0000-0002-6132-0118)

Expecting the Expected: An Analytical Framework to Examine People's Expectations of Robots

We present a novel framework for human-robot interaction designers to analyze and explore expectations of their robot designs. It consists of a model of how people form expectations of robots, and a taxonomy for classifying them. A known challenge in human-robot interactions is expectation discrepancy, in which the expectations people form when interacting with a social robot are not aligned with its actual capabilities. This can disappoint users and hinder interaction. Research has proposed ways to mitigate expectation discrepancy, but designers lack a systematic approach to analyzing and describing expectations. We developed a rigorous theoretical framework by drawing from theories and models from psychology and sociology on expectations between people, and by conducting a field review of expectations in human-robot interactions. We further propose methods for designers to leverage the framework in systematic analysis of how and why people form expectations of a given robot and what those expectations may be. This can empower designers with greater control over people's expectations, enabling them to combat expectation discrepancy.

Keywords: social robotics, human-robot interaction, expectation formation, framework

1 Introduction

Social robots are designed to support collocated interaction with people by leveraging outwardly lifelike social features that people can readily understand [1]. However, 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 predisposition. For example, a person may reasonably assume that if the robot has hands and fingers, then it can pick up items [2]. Of course, the robot may not have this capability, creating an *expectation discrepancy* [2] where people may not only misunderstand how to interact with the robot, but may be surprised and disappointed by a lack of ability, impacting the quality and success of interaction [3]. These misunderstandings can have far-reaching implications including misplaced trust and a host of impacts on how robots integrate into society [4], placing the issue of expectation discrepancy–and managing it–at the center of successful human-robot interaction.

We use the term 'expectations' to refer to a person's beliefs, conscious or otherwise, about a robot's capabilities and potential behaviour. Expectations of social robots emerge from a range of sources, including decades of fanciful media depictions [5,6], and are heavily influenced by the robot's designed form and behaviour [7,8]. Robot designs can align a robot with some known mental category (e.g., an animal) and thus imply capabilities which are commonly associated with that category (e.g., can think, has an emotional system, etc.) [9]. While outwardly human- or animal-like design features may be effective for goals such as promoting familiarity [1] and leveraging empathy [10], they may simultaneously lead to inflated expectations of human- or animal-like capability. As these expectations emerge in part from robot design choices, we may be able to mitigate or avoid inflated expectations, and expectation discrepancy, by designing robots that more accurately imply their capabilities [11,12]. The first step toward this goal of enabling designers to influence user expectations of robots is to better understand this landscape of expectations: how and why people form expectations of robots they encounter, and what kinds of expectations they form.

Continuing the established tradition of consulting work from human interaction to inform approaches to human-robot interaction (e.g., [13–18]), we explored literature on expectations and human-robot interaction, aiming to develop models of human-robot expectations and expectation development. We analyzed key theories 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 *underlying human-robot expectation formation* that unpacks the influencing factors (e.g., robot design, personal experience) and stages that a user goes through to develop, maintain, and update their expectations. Further, we conducted a field review of existing robots, prototypes, behaviors, and literature on expectations of robots more generally, analyzing them for potential expectations and identifying commonalities and salient patterns. This resulted in a novel *two-dimensional taxonomy that describes the range of human expectations of robots*. Together, these two components (expectation formation process model and taxonomy of expectations) provide a novel, comprehensive framework for human-robot expectations.

Finally, we present two new inspection methods for human-robot expectations (*systematic expectation dissection* and *cognitive expectation walkthroughs*) that illustrate how our framework can be leveraged in practice to analyze the expected capabilities of different robot designs. To conclude, we conduct a critical evaluation of our work to identify its effectiveness and limitations and highlight opportunities for future work. Together, these contributions (Figure 1) provide novel tools for analyzing potential human-robot expectations, to support designers in gaining control over expectations and mitigating expectation discrepancies.

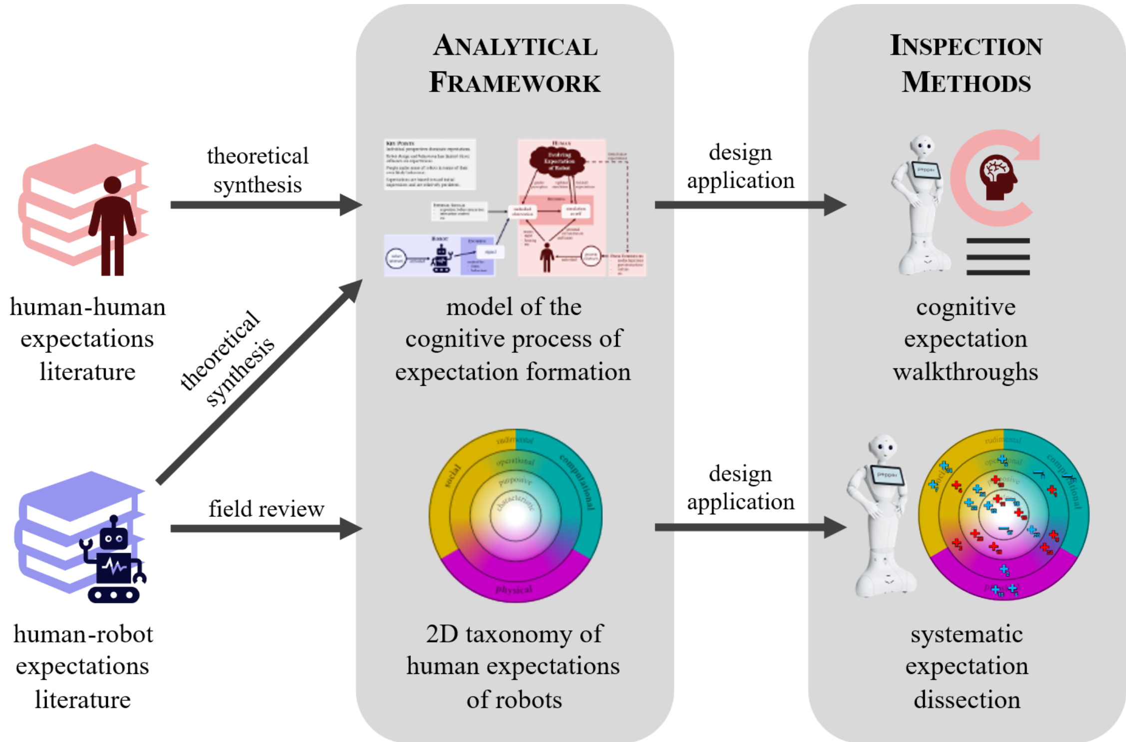


Figure 1. Using literature on expectations in both human-human and human-robot interactions, we produced an analytical framework for examining people's expectations of robots, and proposed inspection methods to employ this framework. These methods are demonstrated in case studies on example robots, including the SoftBank Pepper [102] (pictured), in Appendices A and B.

2 Related Work

A range of existing frameworks describe interactions between humans and robots, including those focused on specific components and properties of an interaction, such as, identifying frequent interaction patterns [19], or classifying aspects such as interaction modalities [20] or basic structural relationships of the participants [21]. Others target specific domains, such as human-robot dialogue, for example classifying instances as linear or branching in nature [14], or identifying fine-grained patterns like repetition [22]. Some frameworks relate to outcomes, such as considering factors that lead people to accept a robot into their homes [13], and many consult peripheral areas to incorporate novel perspectives into the field (e.g., consulting literary analysis for human-robot dialogue systems [14]). In sum, our work builds on this rich methodological tradition of synthesizing knowledge from other fields to provide a human-robot-targeted framework that offers structure and supports analysis of human-robot expectation formation and discrepancy.

2.1 Impact of Robot Design on Interaction

The impact of robot design on expectations, and thus interaction, is well documented, with a large body of work exploring the impact of specific robot design factors such as with respect to users' trust in the robot (e.g., [8,23,24]). Much of this work considers effects of robot aesthetic form, following a common pattern where participants are shown a series of robot variants and asked to rate them on specific metrics [24,25] (sometimes using standardized scales, e.g. [26–28]). One common focus is in linking features to anthropomorphism [29] and how this impacts user reactions [8,30,31], such as the effect of robot sound during movement [32], or the effects of the robot's embodiment (e.g., virtual vs. physical robots) on aspects such as users' trust [23,33].

Similar work looks additionally to the impact of robot behaviors on user expectations (e.g., [7]), including commonly testing the effects of robot mistakes on perceptions and interaction [34–38]. Others have studied the impacts of using social cues [39] such as facial expressions [40], gaze [41], verbal communication [35], and self-gendering [42]. More holistically, a recent work explored using metaphors to explain and understand robots, suggesting that placing a robot into a known, familiar social category can support a person to understand a robot and shape expectations [43]. We complement this growing body of largely-empirical work with a procedural, explanatory perspective on how metaphor and resemblance to known entities can contribute to a person's expectations.

2.2 Expectation Discrepancy

The impacts of expectation discrepancies in human-robot interaction are well documented, often highlighting user disappointment, such as when a person attempts to talk with a robot that cannot converse [44]. These discrepancies can detract from a user's experience [45] and in many cases can create a sense of incompetence and lower trust [46], while a robot exceeding expectations may cause a person to trust and rely on it more [3] (in some cases mistakenly, with potentially dangerous consequences [4]). This relationship between expectation discrepancy and user impressions is more nuanced however, as some robot failures can in fact increase familiarity and likeability [34,36–38]. A more developed understanding of how users form and maintain expectations of robots is thus necessary to determine both how and when expectation discrepancy should be mitigated.

Methods for moderating expectations to be more in line with robot abilities include explaining the capabilities [12], making forms congruent to function [47,48], or having the robot use expressive gestures of incapability [11]. Rosén et al. [18] offers a framework for evaluating human-robot expectation discrepancy that adapts a human expectation formation model [49] identifying a set of factors and metrics to measure expectations and discrepancy. This measures a person's affect toward the robot, expectation of easy interaction, and cognitive load during interaction to identify expectation discrepancies. Our contributions add to this emerging body of work by offering a holistic framework for describing and classifying expectations, and structured tools to systematically analyze how expectations emerge and evolve during interaction.

3 How People Form Expectations of Robots

In this section we analyze current knowledge of how people build expectations of *other people* to inform how we may expect people to make expectations of *robots*. We rely on the assumption that people tend to treat physically embodied robots as if they were alive [50], following concepts of anthropomorphism and zoomorphism (collectively, *animorphism*), the tendency for people to attribute life-like or human traits to non-human entities

[50–54]. Evidence has mounted supporting the fact that people treat robots as lifelike social entities [50] (even more than with other interactive technologies such as personal computers [55,56]) and demonstrating a range of effects including feeling obliged to assist robots [57], engaging them with rapport building behaviors [58], etc.

This tendency may be biological and instinctual, as even infants react to robots as if they were alive [59]. It may also be based in deliberative (conscious) elements [50] or psychological motivations, such as one's need for socialization or potentially inventing social actors (e.g., a social robot) to interact with and rationalize their environment [51]. Regardless, studying how people form expectations of other people can inform how they may form expectations of robots.

3.1 Fundamentals of Forming Expectations Between People

To understand how people form expectations of robots we conducted a literature review of human-human expectations in communication studies, sociology, interaction studies, and cognitive science to identify the dominant theories and models. We did not conduct a full systematic review, which would have mapped out and provided a comparative analysis of the state of the field (including edge cases and open problems) [60,61], but instead aimed simply to leverage relevant aspects of the current understanding of this phenomenon using a standard literature review with a narrative synthesis [60,61], to inform how we may expect people to form expectations of robots. While our exploration and selection contains a qualitative component [60], the result is the identification of four rigorously-developed and well-established theories within their respective communities, serving as critical grounding for our contributions. We present and analyze these below, and synthesize into a cognitive process, to explain how people will form expectations of robots.

This process—drawing inspiration from human-human interaction to inform approaches to human-robot interaction—follows prior successful work, such as using models from social psychology to understand acceptance of robots in homes [13], analyzing human behaviour to inform how robots should act in public spaces [15,16], consulting literature on human conversation to develop frameworks for human robot-dialogue [14,17], or closely related to this work, leveraging a model of expectation development in people to help evaluate expectation discrepancy in interaction with robots [18] (as discussed in Section 2.2).

3.1.1 Message Passing

A predominant paradigm for analyzing inter-personal interaction is message passing [62], which deconstructs complex interaction into a serial set of discrete messages between interlocuters. For example, the *encoding/decoding model* [63] breaks complex communication into a series of messages that are broadcasted by one party (e.g., spoken, facial expressions, gestures, etc., intentional or not) and observed by a receiver (e.g., by listening or watching). All messages go through multiple stages before a receiver can interpret them: messages are encoded, sent (by the sender), transmitted through a medium (e.g., physical world), received, and decoded (by the receiver), before one can make sense of them.

Each phase provides an opportunity for information to be altered, lost, or misconstrued (i.e., corrupted [63]). The observer thus must rely on their particular, imperfect 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* [64]), receiving and decoding a message and developing inaccurate interaction expectations, even when no message was explicitly sent. 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, or *messages* (e.g., a smile, a particular response, that a robot has hands) for targeted analysis regarding expectation formation. Second, we assume that all information is heavily filtered and modified from the transmission and receiving process; these imperfect messages, emitted by a robot, shape expectations.

3.1.2 Expectancy Violations Theory

Expectancy violations theory [65] is a standard lens in communication studies which unpacks interaction between two people. It emphasizes how people hold and maintain expectations of an interlocutor as interaction unfolds or changes. Pre-existing or initial expectations (at the start of an interaction) 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. As the interaction proceeds, new information may not match expectations, creating an *expectancy violation* [66,67]. Violations can be dramatic, such as an expected-to-be calm person becoming surprisingly violent, but are typically more incremental, such as a person taking an unexpectedly informal and familiar tone given a professional situation, or even mundane and unremarkable, such as an unexpected switch in topic within a conversation.

Violations iteratively feed into evolving expectations: new information leads to expectations being revised rather than replaced. Thus expectations are relatively persistent and may be based on pre-conceptions or prior experience [67]. This highlights the importance of prior expectations on interpreting violations. For 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. Instead, if the person introduced themselves as a 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, expectation formation is reflexive: rather than being set according to most recent observations, expectations are the accumulation of ongoing incremental violations over time.

In human-robot interaction, initial expectations may be dominated by predisposition towards technology and prior ideas, often shaped by media portrayals (as argued in [13]), particularly given limited prior experience with robots. These initial expectations are likely to persist even as one interacts with a real robot. Over time, however, we anticipate expectations to evolve incrementally as violations accumulate.

3.1.3 Simulation Theory

Simulation theory provides a complementary view on expectations, postulating that people develop expectations of others by attributing mental states and projecting their own likely behavior [68]), conducting internal cognitive simulations of how *they themselves* would behave given the situation [69]. Mirror neurons may provide biological evidence of this, where neurons activate when observing an action as if one were doing the action themselves [68,70]. In contrast, *theory theory* [sic][1] [71] postulates that people instead systematically apply logical rules or cognitive theories to develop their expectations of how others may behave. Pragmatically we expect people to leverage a combination of simulations and internal theories to develop expectations of others' behavior.

These simulations are necessarily constructed from the observer's individual perspective, biases, and knowledge of the others' circumstances [72], which form a plausible understanding [73]. This explains common problems such as *naïve realism*, where people

---

[1] This is the proper name used in the field to refer to the theory that people use cognitive theories to develop explanations.

see their own experience as an objective reality from which to understand others [74], and *realist bias* [75] or the *curse of knowledge* [76], where a person assumes that their knowledge is shared by others. For example, consider observing someone litter near a clearly visible garbage can. The observer may simulate what would lead *them* to litter [75], perhaps concluding that the litterer has poor moral character [74] (based on their worldview against littering). However, suppose the observer knows the litterer personally and would expect better behaviour. This alternate perspective shapes the simulation, and may instead lead them to hypothesize that the litterer did not notice the trash can, updating their expectation accordingly [72,73]. In either case, simulations are rooted in the perspectives of the observer [71].

Simulation theory has been applied to animals and mechanical devices [77–79], suggesting that anthropomorphism helps people fit observations into existing knowledge to support simulation [80]. Simulation theory supports our position that animorphism leads people to develop lifelike expectations of robots. However, robots present important differences (e.g., robot design, previous knowledge of robots, etc.) that may influence simulations and thus expectations.

3.1.4 Embodied Interaction

Embodied interaction is integral to understanding how people form expectations of others, including robots [81]. From foundations in Heideggerian philosophy, embodiment provides a phenomenological approach to communications studies (e.g., see [82]), and has become central to human-computer interaction under the perspective of *embodied interaction* [81]. Embodiment highlights the role of a person's body and existence within the world (tangible, social, etc.) as foundational to cognition and interaction. All interactions with an 'other'—human, animal, or robot—are mediated through one's embodiment in the world, their *structural coupling* with their environment [83]. In other words, a person's experience of the world (expectations, simulations, interpretations, etc.) cannot be decoupled from their body (size, shape, abilities, senses) and social reality (race, gender identity, nationality, background, etc.).

Embodiment provides a foundation for understanding the critical role of one's own embodiment in message interpretation, expectations violations, and simulation theory. All interpretation is foundationally biased from an individual's own perspective, regardless of the reality of robot capability. 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 exists [84]. We can consider society itself to be constructed from embodied interpretations formed through interactions between people [85].

We analyzed and introduced prevailing theories of how people develop expectations of other people, through passing and interpreting messages, and interpreting the imperfect information to build and refine expectations of others. This includes iteratively updating expectations (through violations) and cognitive simulations of how one would act (simulation theory), all from an individual's particular embodiment. Below we analyze and synthesize these ideas from the perspective of human-robot interaction, developing a cognitive process that explains how a person may form expectations of a robot.

3.2 Synthesis of Human-Human Expectation Formation

We synthesize the above discussion 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, culminating into a *model of the cognitive process of human-robot expectation formation*.

Embodied interaction highlights how people interact with robots, with their personal complex physical and social contexts [86] only narrowly overlapping with the robot's presence within the world [83]. Thus all information received *from* a robot goes through this limited narrow overlap of embodiments, and all observations, messages, and violations are colored by one's biases, world view, etc. (Figure 2); expectations result from observed robot capabilities interpreted within one's embodiment. For example, a robot's cloud computing or facial recognition capabilities are irrelevant if a person cannot observe or understand them (e.g., as with [87]); in this case, the robot's world would include what it can access over the internet, but this online environment is not a part of the person's observed world. We cannot expect people to self-educate, reflect on, or to analyze observations to understand robot abilities. Thus, our first insight into the expectation formation process is that **individual perspectives dominate expectations** more than any objective reality.

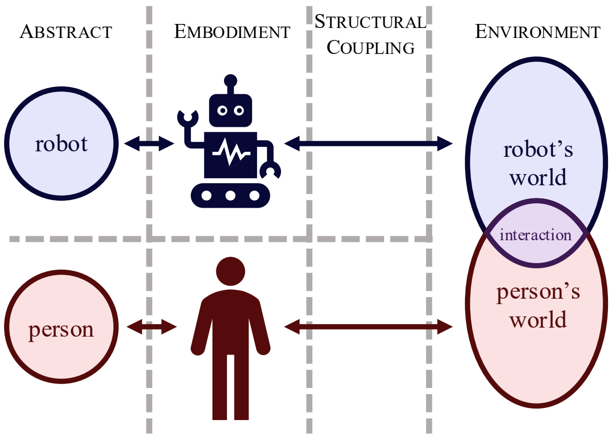


Figure 2. Embodied interaction between a person and a robot where each party is physically and socially embodied and structurally coupled to the world. Interaction between the two parties can only occur at the intersection between their embodiments.

Animorphism, embodied interaction, and simulation theory collectively posit that people perceive social robots and build expectations as if they were alive. People have biological and social tendencies toward animorphism, understand others by simulating their actions [68,70], and have mirror neurons that activate when observing a robot [88–90]. Evidence of this continues to mount for both human-like [88,90] and mechanical designs [89]. Following, we anticipate that people will apply naïve realism [74], projecting their personal circumstances, reasoning, and motivations, onto robots to interpret observations and form expectations (Figure 3). Thus, our next insight is that **people make sense of robots and observations in terms of their own (or another person's) likely behaviour**.

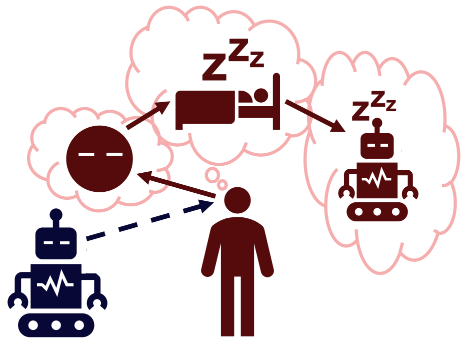


Figure 3. We expect people to make sense of observations using self-simulations based on what they observe. Here, observing a robot with closed eyes, lowered head, and limp arms, a person may simulate themselves, and linking to human sleep, conclude that the robot is in a sleep-like state.

Regardless of designer intent for a robot, the encoding/decoding model [63] highlights that all signals (robot design, behaviours, etc.) are transmitted and translated before being interpreted (Figure 4). As we expect people to draw more from their own understanding than objective robot reality [84], message interpretation may rely on robots as culturally-constructed concepts (e.g., fantastical media depictions) more than as technological objects [91,92]. We cannot expect people to clearly distinguish between fact and fiction for expectations of robots [91]. For example, even if a robot is designed to look stationary by not having legs, a person may apply media-based expectations of robots being mobile and assume the robot has hidden wheels. Thus our next insight is that a **robot's designed features (e.g., visual appearance, behaviors) rely on individual interpretation and only have indirect influence on expectations.** Designers only have limited power to directly shape expectations and should consider designs within the interpretation context.

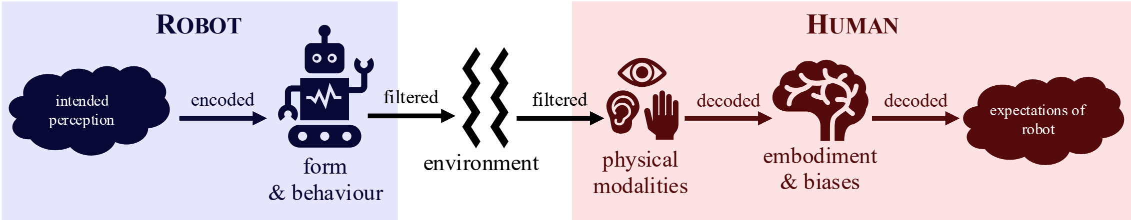


Figure 4. Any objective robot reality is translated and filtered, with many opportunities for alteration and error, and highly biased by the user, before it feeds into building a person's understanding and expectation of the robot.

Simulation theory and embodiment both highlight how an individual's predisposition and background influence how they interpret received information—tendencies that are likely to resist change, even when confronted with new information [72]. Consequently, new information is processed within one's embodiment and predispositions to *update* existing (perhaps prior) expectations. Expectations tend to be resistant to change, and violations rarely lead to entirely new expectations; rather, they evolve reflexively [67]. It generally requires accumulated violations to greatly alter expectations, even quickly-adopted first impressions [93]. For example, empirical evidence in HRI has demonstrated the lasting effect of first impressions [94], and how impressions evolve with repeated interactions [95] (Figure 5). This underscores the importance of understanding a person's background when predicting how they will interpret robot designs. For example, the *pratfall effect* demonstrates how a person may be seen as more likable when they make mistakes, *if* they were initially seen as competent [96]. Conversely, a person previously seen as incompetent may be seen as less likable upon making the same mistake, a result also observed in human-robot interactions [36]. This results in the insight that **expectations are biased toward initial impressions, relatively resistant to change, and are reflexively *updated* with new information, rather than being replaced**.

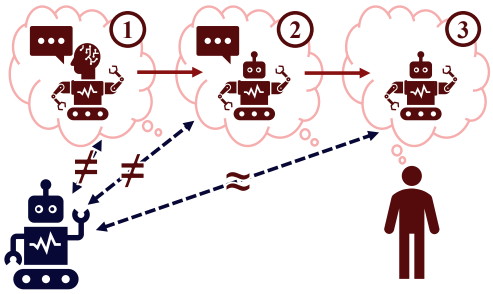


Figure 5. Expectations evolve during interaction, starting from a-priori beliefs; new information *modifies* existing expectations. For example, an observer (1) upon seeing a humanoid robot may assume intelligent interaction ability. (2) Poor conversation behavior may lower expectations but the person may still assume it can talk. (3) Only after observing continued poor ability do they perhaps expect it cannot talk at all. Initial expectations thus change gradually, and can be recalcitrant in the face of contrary evidence.

All these insights emphasize the significant conceptual gap between the objective reality of a robot's capabilities and actions, and expectations that people form about the robot, with many steps of indirection, translation, and interpretation.

3.3 Model of the Cognitive Process of Human-Robot Expectation Formation

The four key insights presented in Section 3.2 each describe an important component of human-robot expectation formation. Shifting our focus more broadly, we further synthesize these points into a detailed process that describes and analyzes how we expect a person to develop and maintain expectations of a robot they encounter. This *cognitive process model* encapsulates several simultaneous processes and inputs that shape a person's evolving expectations of a robot.

For illustrative purposes we detail a potential pathway through the inherently parallel process (Figure 6). A robot emits visual and behavioral design signals, which are encoded through its embodied form as they are transmitted into the environment. Alongside these are peripheral signals that provide exposition such as introducing the robot or the context of use (e.g., a factory), which may not relate to objective robot capabilities. A person receives these signals, interpreting and processing from within their biased physical and social embodiment. All of these signals together become inputs into the person's internal cognitive processes of expectation formation.

The person simulates what their observations would mean for them, promoting animorphic interpretation. This feeds into evolving robot expectations, weighted toward persistent predispositions and prior expectations. Observations feed back into long-term experiences and become prior expectations that over time influence new ones.

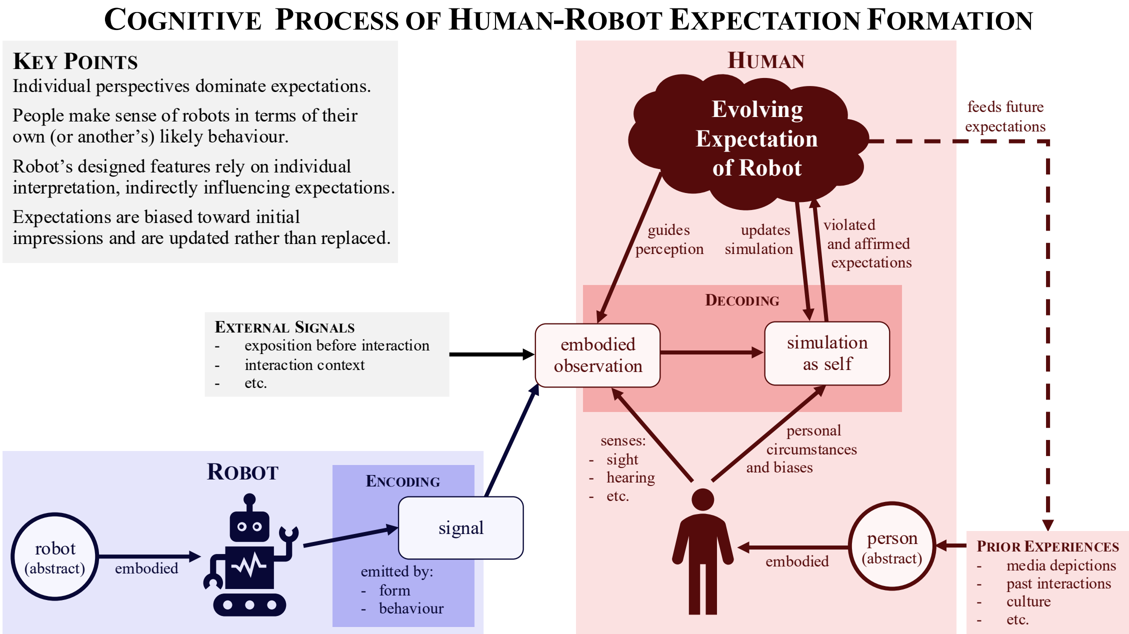


Figure 6. Our proposed Cognitive Process of Human-Robot Expectation Formation illustrating how people form and maintain expectations of robots.

Our cognitive process model helps unpack the significant gap between robot design, objective capabilities, and resulting expectations, as a concrete analysis tool for exploring expectations and anticipating results of hypothetical robot designs. Designers can thus use this model to support efforts of mitigating expectation discrepancy while acknowledging limits of designer influence. In the following section, we employ this model as an exploratory guide to consider the range of possible expectations that people may form of robots.

4 Classifying Expectations: A Novel Taxonomy

So far, this paper has followed the trend of using 'expectations' less formally to refer to general beliefs in the presence or absence of robot qualities and capabilities (e.g., [2,7,45]). However, designers need more precise language to specify and differentiate expectations, for example, that a robot can physically move or be a friend. While some work provides specific definitions, such as using future event probabilities [18], these are limited to targeted instances. Instead, we develop a broad taxonomy that covers the wide range of expectations regarding a robot or interactions with it.

4.1 Development Process

Through our background investigation into human-robot expectations, we found no existing literature enumerating or quantifying the types of expectations that people may develop. Thus, we conducted a semiformal field review to generate a representative expectation dataset with broad coverage which we could inductively analyse for dominant themes, focusing on saturation of expectations discovered. We did not conduct a systematic review as our process does not require a full nuanced summary on the state of knowledge [97]. While a systematic review would more exhaustively identify possible expectations, our goal of identifying the key, but broad themes and categories of expectations does not require a full enumeration. However, we highlight that our resulting taxonomy can serve as structure for more exhaustive systematic reviews.

We collected research literature to represent recent results and the state of expectation discourse, as well as robot platform and behavior design exemplars (both real and science fiction) to serve as data for cognitive process analysis (Section 3). We searched via Google, Google Scholar, and the ACM Digital Library using the keywords 'robot', 'expectation', 'impression', and 'evaluation'. This, resulted in a corpus of research papers, images, videos, and behaviors representing a broad range of expectations and potential design elements, that we aim to cover with our taxonomy.

Our next step was to analyze this corpus to enumerate a range of possible expectations that people may form, using literature, analysis and informal brainstorming supported by our cognitive process model (Section 3). This process of open-ended ideation of expectations was conducted following standard HCI principles to explore potential user experiences [98]. Using real data from expectations we found in the literature as a starting point, we expanded our list using by considering how these expectations may evolve and change over the course of an interaction based on our model of expectation formation. For example, starting from the expectation that a robot can talk (a simple expectation which is well-attested in the literature), we considered interactions in which the user has this expectation violated to examine how they may interpret the robot's silence. As our model anticipates that the user will understand the silence in terms of their own behaviour, they may conclude that the robot is refusing to speak to them, and develop a new expectation that the robot is aloof, or perhaps hard of hearing, depending on the context and their past experiences. Thus our model served as a generative exploration tool to aid in the generation of a broad, plausible corpus of expectations of robots. This resulted in an expansive list of plausible expectations with wide coverage of robot designs.

Following, we employed standard inductive, iterative thematic analysis to distill these expectations into a minimal set of dominant categories [99], supported by informal affinity diagramming (e.g. [100]). We started with initial intuitive categories, iteratively redefining and reclassifying expectations until our categories succinctly and accurately described our corpus. As is common with qualitative methods, we note that this process involves a degree of subjectivity that is embedded in the resulting classification [60,101];

however, this process results in a novel and robust perspective to support a deeper understanding of people's expectations.

4.2 Taxonomy of Expectations

Based on the salient themes identified in our review, we constructed a two-dimensional taxonomy: one dimension categorizes expectations based on the type of capability (e.g., physical, social, etc.) and the other categorizes based on the level of abstraction (e.g., motor abilities vs. attributed personalities). For brevity we describe these categories with examples involving the SoftBank Pepper [102] and Sony aibo robots [103] (Figure 7) instead of extensive data from our corpus.

Figure 7. The SoftBank Pepper [102] (left) and Sony aibo [103] (centre) used as examples throughout this section.

4.3 Taxonomy Dimension: Domains of Expected Capability

We identified three primary groupings of expectations of robot capabilities (Figure 8):

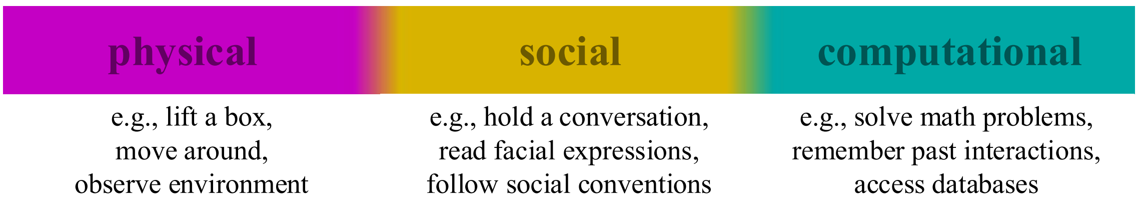


Figure 8. Examples of expectations in each of the three domains of expected capability. The blurred domain boundaries indicate that some expectations may span multiple domains.

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 actuators, including manipulators or wheels, and the strength, fine motor, or general movement or locomotion ability, etc. of those actuators. For example, people may expect that Pepper can use its arms to wave at them, or the legged aibo robot to walk across a room. This includes outputs such as light or sound emissions, and sensory capacities such as being able to see, feel, touch, or receive radio transmission.

Social Capabilities – People form expectations about a robot's social abilities, including communication and participation in society. For example, they may expect that the humanoid Pepper can speak, hold a conversation, use social gestures (such as a wave or high five), or pay attention to a person. Similarly, people may expect the dog-like aibo to have internal emotional states and to be able to interpret theirs to some degree. People may further expect a robot to understand interpersonal relationships, social dynamics in a group, or participate in social conventions such as yielding access to an elevator when appropriate [104].

Computational Capabilities – People form expectations about a robot's ability to perform computation, encompassing a similar range of expectations to those of a traditional PC. People may expect that Pepper can perform mathematical or logical calculation, or that aibo can remember their face and past interactions with inerrant, computer-like precision. This can include access to information sources (e.g., databases, encyclopedias, etc.), or learning capability.

Note the blurred boundaries between categories in Figure 8; this indicates that expectations can span categories. For example, the expectation that aibo can learn to perform tricks relates to both physical abilities (perform movements) and computational abilities (learn and remember tricks), and the belief that Pepper will shake a person's hand is both physical and social.

4.4 Taxonomy Dimension: Levels of Expectation Abstraction

We found expectations to range from purely mechanical (e.g., a motor can move [105]) to high-level complex behaviors such as intentions and personalities (e.g., [106]). Our analysis resulted in four ordinal abstraction levels (Figure 9).

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 9. Examples of expectations representing each of the four levels of expectation abstraction.

Rudimental Expectation – People form expectations of basic mechanical robot capabilities, independent of the robot's environment. For example, people may anticipate that Pepper can speak and perform calculations, or that aibo's legs have motors sufficiently strong to walk. This is expectation of raw capability, not a robot's ability to use it to perform operations.

Operational Expectation – People form expectations that a robot can use its rudimental capabilities (e.g., has motors, can calculate) to perform specific operations in its environment (e.g. can lift a box, can solve a math problem). This covers what a robot *could* do in practice; for example, a person may expect that Pepper can engage in friendly conversation (given speech ability), or that aibo could climb over a small obstacle (given its legs).

Purposive Expectation – People will form expectations of a robot's goals, what it intends to do, using its operational capabilities. For example, given that a person believes a warehouse robot can lift a box, they may expect that it will try to collect boxes. Conversely, operational capability and intention may not align; a person may believe that Pepper will not idly chat with them in a busy, task-focused context, despite having that ability. Expectations of high-level goals can shape expectations of intended actions; if an aibo aims to navigate across the room, a person may expect it to climb over obstacles in its way. Purposive expectations can further draw from animorphic attributions of will and desire, such as believing that aibo—analogous to a dog—wants attention from the user and will act accordingly.

Characteristic Expectation –People will form expectations of a robot's characteristic behavior, analogous to a personality and similar to animals or other people. For example, one may assess their particular aibo as being strong and reliable or a Pepper they encounter as being friendly but professional. These traits can be more or less animorphic in nature (e.g., 'strong' may simply refer to an overall assessment of the robot's mechanical strength, while 'friendly' ascribes a life-like personality to the machine).

Note how Figure 9 has clear boundaries, in contrast to the blurring in Figure 8: all expectations in our corpus could be cleanly placed into a single expectation abstraction category. However, we found considerable dependency between the layers of abstraction. For example, if a person expects that a robot has eyes and can see (rudimental), it may be natural to assume that it can recognize people (operational) and is trying to monitor them (purposive). Inversely, if a person expects a robot to be chatty (characteristic), this may infer lower-level expectations such as that the robot wants to talk to them (purposive), is able to hold a conversation (operational), and has speakers and a microphone (rudimental). Expectations resulting from this logic may not match robot capability, resulting in expectation discrepancy.

4.5 A Two-Dimensional Taxonomy of Expectations

Together, the domains of expected capability and levels of abstraction form two orthogonal dimensions of a taxonomy of expectations one may develop for social robots, enabling us to categorize and position how expectations relate to one another. Any given expectation has both a capability domain (physical, social, computational) and a level of abstraction (mechanical, personal, etc.).

We propose a polar two-dimensional diagram to visualize the taxonomy, plotting domains of expected capability along the angular axis (around the circle) and levels of expectation abstraction along the radial axis (from the outside inward to the centre, Figure 10). Domain is simple to differentiate at the rudimental level, for example, with the ability to move (physical), calculate (computational), or talk (social). However, this clear binning is more challenging at higher levels of abstraction; for example the operational ability of giving a hug has physical and social components, or a robot being 'greedy' may have more vague social and computational components. We visualize this transition in Figure 10 by having both clearly divided regions at the outer layer, gradually blurring together toward the more abstract core. Using this visualization we could imagine plotting specific expectations within the space, which we will explore in the following section.

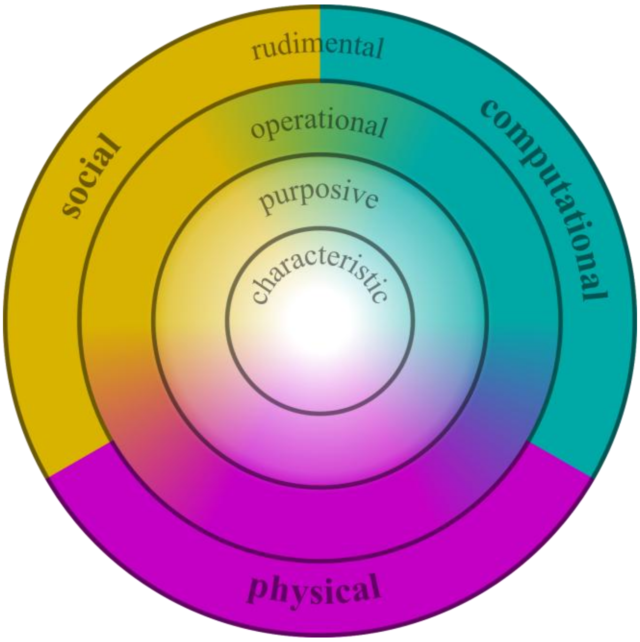


Figure 10. A visualization of our two-dimensional taxonomy of expectations of robots, with capability domains on the angular dimension and levels of abstraction on the radial dimension. Note that the line between the capability domains blur as one moves further away from rudimental capabilities, as the deeper, more abstract 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 in Section 5.

We emphasize that the blurred, continuous depiction of the domain dimension is not meant to imply a precise scalar quantity, but rather an approximation to represent the way that more abstract expectations can straddle across dimensions. A polar diagram was chosen to represent this taxonomic space because it highlights these increasingly blurred boundaries. While we explored other arrangements that may be simpler at first glance, including a rectangular matrix, we found that the inability to coherently display all domain boundaries made it difficult to highlight the proximity of the domains on the edges of the diagram.

5 Sample Applications: Inspection Methods

In this section we aim to bridge the gap between the theory and practical application of our framework by providing analytical tools using our framework that support researchers and designers in engaging with the problem of expectation discrepancy. Drawing from HCI analytical evaluation methodologies [107], we designed two analytical techniques for exploring expectations: *systematic expectation dissection*, to identify areas of expectation discrepancy, and *cognitive expectation walkthroughs*, to support explanation and understanding of identified discrepancies. We present these methodologies below, with full case-study applications provided in Appendices A and B.

We note that these methods serve as a component of our evaluation: the sample applications provide an illustrative evaluation that highlights how our framework can be used to focus a designer's attention and guide the exploratory process. This follows established practice in human-computer interaction [108], particularly with theoretical frameworks such as this [14,20,62], where evaluation of a toolkit's potential is provided through concrete demonstration of the framework's application.

5.1 Systematic Expectation Dissection

We propose *systematic expectation dissection* as a novel methodology for leveraging our framework to analyze observed or predicted user expectations of a robot. This is an exploratory process that guides a designer to systematically explore potential user expectations and discrepancies within our full taxonomy space. Designers plot results on a simple visualization to organize them, identify trends or blind spots, and to communicate results to others.

5.1.1 Visual Expectation Plotting on the 2D Taxonomy

We can plot expectations within the two-dimensional visualization of the taxonomy as presented in Figure 10. As the taxonomy is not scalar, but rather nominal (capability domains) and ordinal (levels of abstraction), we plot within general regions of the visualization only and are not concerned with exact coordinates.

When plotting an expectation, we denote whether the person expects the robot to have or not have an ability, which we call *polarity:* we plot a **+** to indicate that a robot has a feature (e.g., it *can* talk) and a **–** to indicate that the robot does not have a feature (e.g., it *cannot* walk). Finally, we include the relationship of the expectation to the robot's capabilities, for example, an accurate expectation or a discrepancy. We represent this using color, with blue representing accurate *matching* (e.g., +, –) and red indicating a discrepancy (e.g., +,–). For example, a correct belief that a robot cannot walk is a matched, negative expectation (–) while a mistaken belief that a robot can speak is a positive but discrepant expectation (+) (shown in Figure 11).

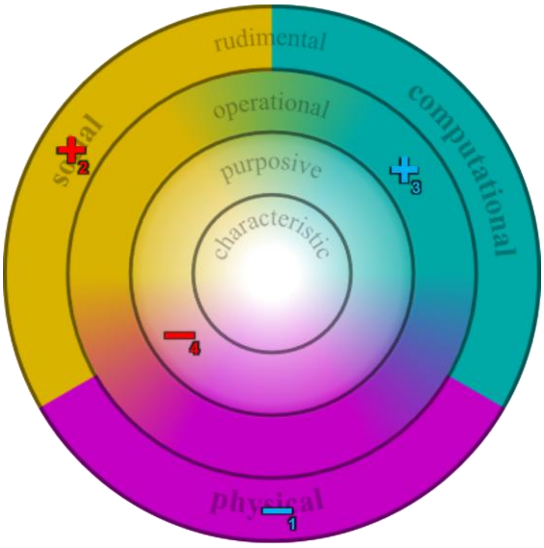


Figure 11. We can plot expectations into the taxonomy space using the icons explained in Section 5.1.1. Expectation 1 (–) represents a correct belief that a robot cannot walk, 2 (+) is a mistaken belief that a robot can speak, 3 (+) is a correct belief that a robot can solve a given math problem, and 4 (–) is a mistaken belief that a robot won't want to shake a user's hand.

We illustrate these as binary for simplicity while acknowledging that a given expectation may not be simply matched or mismatched, but rather more nuanced. However, our coarse-grained classification supports exploration and simple visualization.

5.1.2 Systematic Expectation Dissection Procedure

One starts a *systematic expectation dissection* by compiling a list of expectations one may hold about their robot. This could result from exploratory user studies on the robot, analogizing from studies of similar robots or relevant literature, or critical analysis of the design. Depending on the context of interest, this list may be focused on expectations at a particular point in the interaction, such looking specifically at initial impressions, or at expectations developed after extensive interaction. The objective in this step is to collect sufficient data to support a designer or researcher in engaging with the full range of potential expectations.

Following, one classifies each expectation regarding the *polarity* and *matching* (e.g., using +, –, +, –) plots, and starts to map them onto the taxonomy space. Each expectation is first classified into a level of abstraction, and as appropriate, assigned to a capability domain. For overlapping domains an expectation can be closer to or on a boundary, or placed more into the blurred regions. This process will produce an aggregate graphical summary of potential expectations that highlights both polarity as well as discrepancies. Designers can examine this summary to identify patterns, for example concentrated areas of matched or mismatched expectations, which may suggest strengths and weaknesses of the robot's design.

To assist with applying this technique Appendix C includes a printable visualization template (Figure 10) and the expectation process model (Figure 6) for reference. We envision that a designer may print this sheet and use it to manually plot user expectations of their robot.

5.1.3 Systematic Expectation Dissection Case Studies

We performed four case studies, applying our systematic expectation dissection technique to four robots, for demonstration and informal evaluation purposes. We analyzed the SoftBank Pepper [102], Boston Dynamics' Atlas [109], the Sony aibo [103], and SnuggleBot [110]. We selected these as representative of dominant morphological categories of social robots: Pepper and Atlas are humanoid, approximately human-sized robots; aibo is a pet-inspired zoomorphic robot; and SnuggleBot is a 'cuddly' companion robot (similar to Paro [111] and LOVOT [112]). This selection serves to illustrate how the technique can be applied across a diverse selection of robot designs, with two robots of similar form (Pepper and Atlas) included to show how it can be used to highlight the effects of smaller changes in design.

To conduct these case studies, we generated dummy expectation data rooted in our field review of expectations (Section 4) to serve as a sample input to the process and to demonstrate the results of a systematic expectation dissection and to support analysis. We present these visualizations shown in Figures 12-15. We present the full details of these case studies (including the enumeration of the sample expectations) in Appendix A.

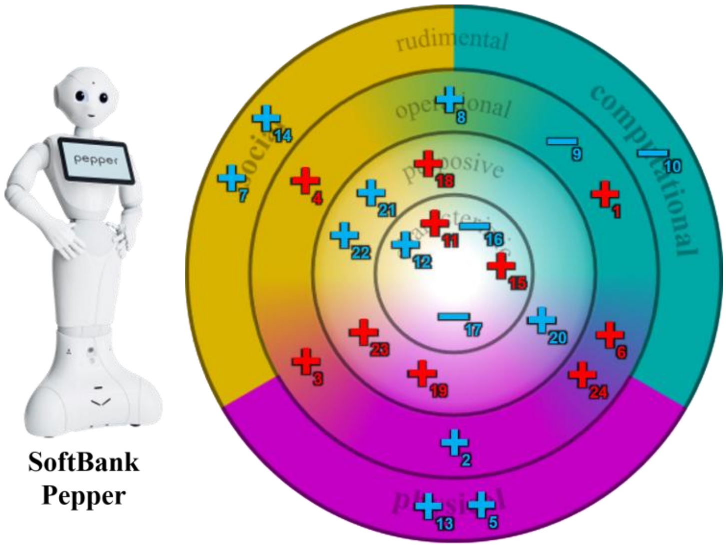


Figure 12. Example expectations of the SoftBank Pepper [102] visualized with our expectations taxonomy.

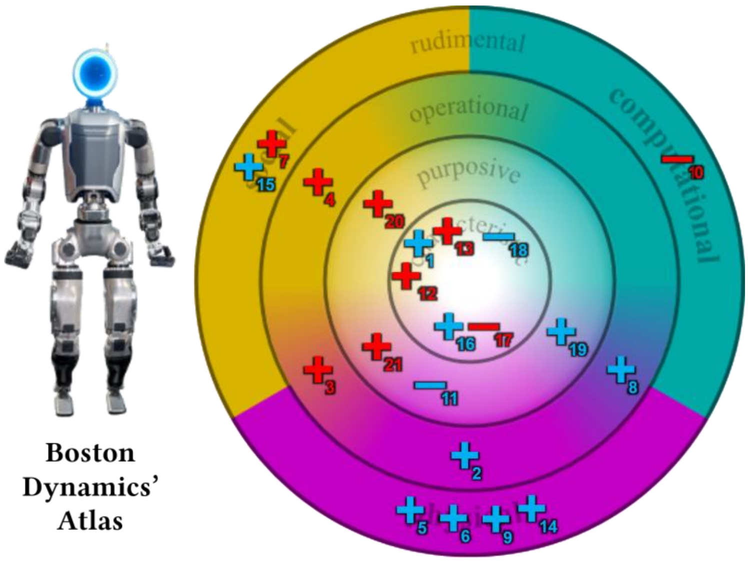


Figure 13. Example expectations of the Atlas [109] visualized with our expectations taxonomy.

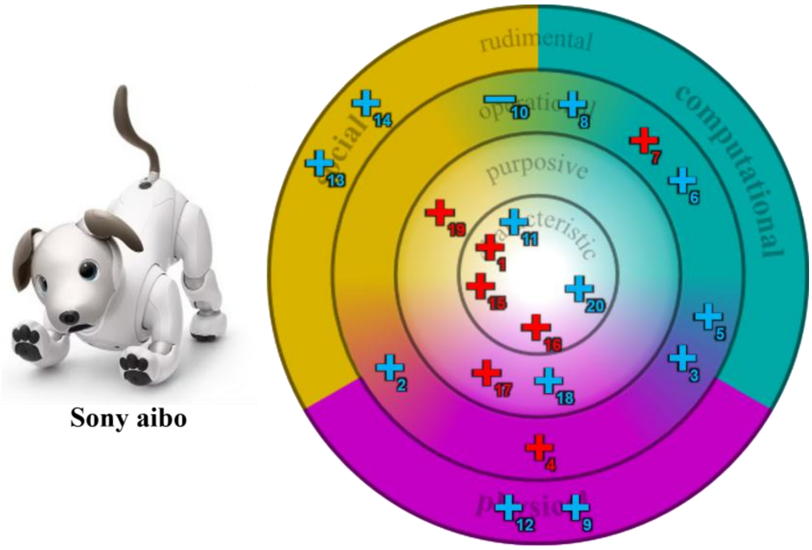


Figure 14. Example expectations of the Sony aibo [103] visualized with our expectations taxonomy.

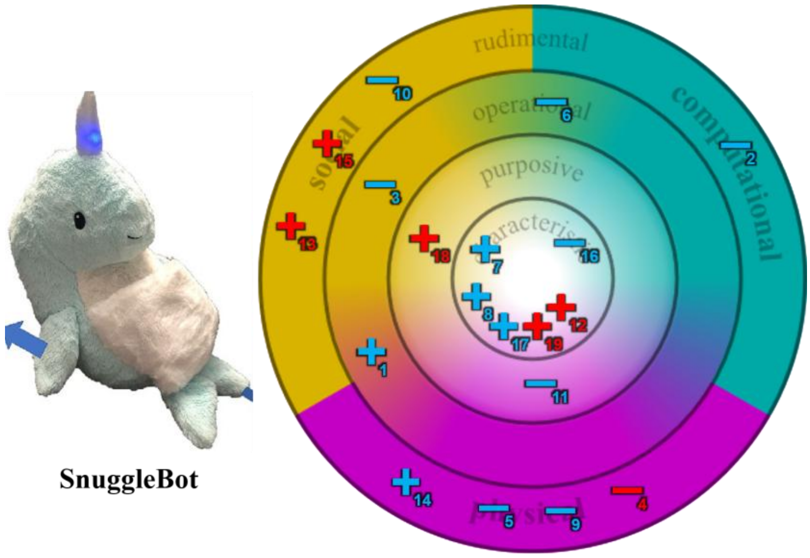


Figure 15. Example expectations of the SnuggleBot [110] visualized with our expectations taxonomy.

These visualizations highlight clusters of expectations and support comparison. For example, perhaps SnuggleBot may not generate as much computational expectation as the other robots, and Pepper may be more 'balanced' across the taxonomy. Further, this highlights clusters of expectation discrepancy (red plots); for example, SnuggleBot may have more social discrepancies while Pepper may generate discrepancies across the whole taxonomy. Despite their similar forms, Atlas may generate more physical expectations than Pepper, while maintaining similar but more discrepant social expectations, such as being expected to speak when it cannot. Such visualizations and comparisons across robots may aid designers in developing an understanding of how robot designs influence expectations (and in particular, what kinds of expectations), and may also be employed between particular revisions of a single design to compare the effects of smaller adjustments. Given sufficient empirical data from users, this technique may be used to display and summarize definitive differences in expected capability. Although these observations are based on dummy data, they demonstrate the utility of the systematic expectation dissection technique. Thus this technique can support researchers and designers to comprehensively explore the full range of robot expectations and discrepancies.

5.2 Cognitive Expectation Walkthroughs

We propose *cognitive expectation walkthroughs* as a scenario-based analytical technique for exploring the process of robot expectation formation. Following established HCI methodology [107], a cognitive expectation walkthrough centers around establishing personas and tasks or scenarios, and step-wise following through the scenario to analyze interaction. In this case, at each step an evaluator (or group) would apply our expectation formation process (Section 3, Figure 6) to explore how expectations may form and evolve as interaction unfolds. This method provides a cognitive framing for understanding how a person may develop and maintain expectations of a robot, providing insights on how designs may be altered to mitigate expectation discrepancies.

5.2.1 Cognitive Expectation Walkthrough Procedure

A cognitive expectation walkthrough requires three key components: a robot platform and behaviour description, user personas, and a scenario. The method requires a clear and detailed robot design (including physical, visual, behavioural, etc., features), as these are central to expectations; robot and implementation-agnostic walkthroughs would fail to account for the pivotal role of a robot's design on expectations. Further, at least one detailed persona, including the individual's background, biases, and other details relevant to shaping expectations, and a clear scenario or task involving the robot, are essential for providing the key context and goals that drive the interaction.

Determining the scenario and user personas is an integral part of a cognitive expectation walkthrough. These choices inform the external signals, prior user experience, and initial expectations, which serve as essential inputs to the expectation formation process. As there will typically not be a single appropriate choice for this background information, it is necessary to carefully consider the contexts in which the robot may be deployed, and may be helpful to conduct multiple walkthroughs with different scenarios and personas in order to cover a broader range of potential interactions. Unfamiliar readers can consult human-computer interaction texts for details on persona and scenario development (e.g., [107]).

Given these components, an evaluator can conduct the cognitive expectation walkthrough by considering, step by step, how the interaction may unfold for the given robot, personas, and scenario. At each step, an evaluator applies the cognitive process model to explore how expectations develop: this systematically considers all signals received (including robot form and behavior, environment, etc.), tracing them through the cognitive process to evaluate how they impact iteratively evolving expectations. The result is a rich description of what expectations may be anticipated, how they may emerge and evolve, and how this relates to the robot's design and person's background.

5.2.2 Cognitive Expectation Walkthrough Case Study

We executed an example cognitive expectation walkthrough for a scenario where a store customer approaches a SoftBank Pepper robot [102] that is programmed and presented as a shopping assistant. We developed detailed hypothetical scenarios, a persona, and robot behavior, and step-wise evaluated these using the cognitive process model (Figure 6). We present our results here, and provide full details in Appendix B that interested readers can examine for more detail.

The results of our analysis highlighted how the robot's visual humanoid design, its behavior design of using friendly verbal greetings and human-like gestures, and behavior of maintaining gaze with the person, are all anticipated to promote expectations of advanced conversational capability. This is further reinforced if the person is a tech enthusiast with related media exposure (as may be expected of a person approaching a robot). This highlights the inevitability of a person naturally expecting that this robot could hold a smooth conversation. Thus, unless such a robot has very sophisticated social conversational ability we could anticipate expectation discrepancy. Further, our exploration highlighted how initial expectations may not change as a person observes failures, and we can expect repeated failures to be needed for a person to finally understand the robot's limitations.

Thus, our cognitive expectation walkthrough clearly highlights the challenge with creating a retail or similar kiosk using the Pepper robot and common behavior implementations, without resulting in over-inflated expectations and ultimately expectation discrepancy and user disappointment. In general, this method is a potential tool for supporting one to engage intricately with a potential or real robot design. By leveraging our cognitive process model, this method encourages evaluators to consider a broad range of factors (robot design, person's background, and tendencies toward expectation development) that may contribute to resulting expectations and related behaviors.

5.3 Analytical Techniques for Expectations of Robots

We presented two techniques for systematically exploring a person's expectations of a robot design. *Systematic expectation dissection* allows for comprehensive analysis of expectations across the expectation taxonomy space, providing simple, bird's-eye view visualizations that support meta-analysis and comparison of expectations between robot designs, and allowing designers to explore how design variants and may lead to different expectations and expectation discrepancies. *Cognitive expectation walkthroughs* allow for systematically analysis of how expectations may form and evolve, based on knowledge of human expectation formation, given a robot design and scenario. Together, these techniques illustrate the potential utility of our framework (cognitive process model of human-robot expectation formation and expectation taxonomy) for describing and examining expectations of robots, toward empowering designers to mitigate unwanted expectation discrepancies.

6 Evaluation and Critical Reflections

Given that there is as-of-yet no comparable encompassing framework on robot expectations that we can compare against (e.g., as a baseline in a study), direct quantitative evaluation is challenging. Further, studies with potential robot designers (e.g., workshops in our lab) would not have sufficient ecological validity given our limited access to the HRI expert who would use our work, and therefore appropriately evaluating the pragmatic utility to robot designers will require consideration of the work's use after publication.

Therefore, following established practice developing frameworks in human-robot interaction (e.g., [13,14,18,19]), the primary evaluation of our work is inherent in our theory-driven integrative approach. That is, the synthesis of literature and the model's alignment and agreement between diverse theoretical constructs, grounded in well-developed theories and background work, provides a critical grounding for the resulting framework.

Additionally, in Section 5 we presented analytical evaluation techniques [113] as illustrative examples of how our framework may be employed in practice. Drawing from descriptive evaluation methods, such as in design research [114], this approach uses reasoned arguments and scenarios to demonstrate the internal logical consistency and showcase the utility of the framework for focusing attention and guiding exploration and reasoning through expectation formation in human-robot interaction.

To further strengthen this evaluation and support designers in understanding how they may employ this work, we conduct a critical reflection on our framework and proposed inspection methods, based on our experiences developing (Sections 3 and 4) and applying (Section 5) them. We first situate our framework within the context of prior approaches to engaging with expectations in human-robot interaction, comparing it against earlier frameworks and considering how it may complement them. Following, we detail the strengths and limitations of our framework in supporting a more detailed understanding of user expectations of robots.

6.1 Comparison With Other Frameworks

As part of our evaluation we explicitly compare and contrast our framework against existing work in the space, which we use as a baseline for the current state of the art in exploring expectations and expectation discrepancy. No other framework provides direct overlap with ours in terms of the objective of the tool and the questions answered by the output, which precludes us from making direct comparisons of their effectiveness. Instead, our evaluation below considers how our framework can complement and fit within the broader context of tools and approaches, and how it offers novel perspectives. Specifically, we compare against the following literature identified in Section 2: Rosen et al.'s [18] Social Robot Expectation Gap Evaluation Framework, Dennler et al.'s [43]

Design Metaphors for Understanding User Expectations, as well as scale instruments for measuring user perceptions of robots (e.g., Godspeed [26], RoSAS [27], NARS and RAS [28], etc.).

Rosén et al. [18]'s Social Robot Expectation Gap Evaluation Framework provides a set of metrics that can be used to quantitatively evaluate a person's degree of expectation discrepancy toward a robot. This framework requires a practitioner to collect significant data including participant questionnaire and interview responses, as well as quantitative data such as interaction duration and reaction time. The goal in this framework is to measure peoples' affect toward the robot, cognitive load, and expectations of interaction ease. Rosén et al [18] then proposes analysis of this data to quantitatively evaluate expectation discrepancy severity and direction (i.e., the robot exceeded or fell short of expectations), although the work does not offer a detailed method to perform this analysis. It also does not provide mechanisms to support an evaluator to more broadly explore what kinds of expectations may emerge, or what may have caused them. In contrast to this, our framework provides consistent vocabulary and a structured approach for investigating these discrepancies and analyzing the expectations at a more granular level, without necessarily requiring large quantities of empirical data. Thus, we envision that this framework may be used as a starting point to identify instances of discrepancy, which can then be analysed more deeply using our own framework.

Dennler et al.[43]'s Design Metaphors for Understanding User Expectations presents an approach to explaining expectations by considering how the user may understand a robot via application of a metaphor to a more familiar category. For example, a highly anthropomorphic robot like Pepper [102] may be understood by metaphor to a person, encouraging expectations of advanced human-like abilities, while a robot like SnuggleBot [115] may be understood by metaphor to a doll or toy, suggesting limited interaction or intelligence. This approach provides a simple-yet-powerful mechanism for understanding what expectations people may have of a robot for specific cases that fit into clear metaphors; however, they do not provide a method for unpacking these expectations, vocabulary for explaining them, or a method to understand the process of developing them as provided by our framework. As such, this approach is complementary to our own, where these metaphors could fit into the 'prior experiences' component of expectation formation (Figure 6).

Several scale instruments exist for measuring perceptions and thus perhaps expectations of robots (e.g., Godspeed [26], RoSAS [27], NARS and RAS [28], etc.). These tools generally summarize a person's attitudes toward a robot along several dimensions (commonly including intelligence, warmth, animacy, and competence, among others). When considering these tools in comparison to our own taxonomy, we notice that they primarily focus on characteristic expectations (e.g., that Pepper is warm and competent), clustered within the innermost, abstract layer of our taxonomy. Our taxonomy highlights the potential for the extension of this scale approach to measure other forms of expectations (e.g., more rudimental or operational). Further, the output of these scales may be used as one source of expectation data to be analysed through our systematic expectation dissection technique.

Table 1 presents a summary of our own framework and the above approaches, highlighting the unique perspective our framework provides and how it may complement these prior perspectives. Our framework is unique in providing a visualization of user expectations which captures a broad multidimensional view of expectations, combined with an analytical approach for examining and explaining discrepancies. It does not replace any existing approach, but rather serves to be employed in concert with them. The new perspectives offered by our framework thus support designers in developing a deeper understanding of user expectations of their robots.

	Our Framework		Prior Approaches		
	Taxonomy, with systematic expectation dissection (Sections 4, 5.1)	*Process model, with cognitive expectation walkthrough (Sections 3, 5.2)*	*Social Robot Expectation Gap Evaluation Framework* [18]	*Design Metaphors for Understanding User Expectations* [43]	*Perception measuring instruments (e.g., Godspeed* [26], *etc.)*
General function	Explore and visualize the range of expectations a user may hold	Support structured analysis of how and why a person may form expectations of a robot	Evaluate the level of expectation discrepancy a user experiences when interacting with a robot	Emphasizes how expectations of robots can be influenced by robot's design similarity to a more familiar category	Quantitatively summarize a user's attitude toward a robot
Inputs	Data on user expectations	Specified robot, user, and interaction scenario	Questionnaire and interaction data	Robot design	Questionnaire data
Outputs	Detailed, visual summary of expectations to support comparison and analysis of trends	Insights into what inputs and formation steps may be leading to the development of particular expectations	Evaluation of expectation discrepancy, broken down into severity and direction	Relevant design metaphors that hint at potential expectations of capability and behaviour	Quantitative ratings of robots along various dimensions
Example: Potential output with Pepper case study scenario (Section 5.2)	Visualization showing Pepper creates varied expectation discrepancies across domain and abstraction (Section 5.1.3)	Walkthrough of an interaction with Pepper highlighting how certain features may lead to discrepancy (Section 5.2)	User experienced severe disconfirmation of their initially high expectations of Pepper	Pepper's human-like appearances lead to expectations of advanced, human-like capabilities	Pepper is warm and competent.
Relationship with our framework			This framework may be used to detect expectation discrepancy which can be further analysed with our framework	This work expands on familiar categories as a major input into our expectation formation process (within the 'prior experiences' component)	Our framework highlights potential to extend scale concept to less abstract expectations Output from these scales may work as input for systematic expectation dissection

Table 1. A summary of the approach and usage of our own framework in comparison to prior perspectives for engaging with human-robot expectations. We summarize the overall function of each tool, the typical inputs it takes and outputs it produces, an example of what that output may look like if applied to our case study scenario in Section 5.2, and how it may be employed in concert with our framework.

6.2 Reflections and Limitations

By reflecting on the process of developing and applying our framework, we have identified key strengths and limitations in its scope and perspective, which we detail below.

6.2.1 Taxonomy Scope and Granularity

Our taxonomy was able to easily cover the full range of expectations compiled through our review of prior work and ideation process; while we cannot say whether this taxonomy exhaustively covers all possible expectations, this shows that its scope encompasses the full range of common expectations we encountered in our review. As an initial framework, our objective was not to propose a singular, definitive classification scheme that captures some underlying structure of all expectations, but to offer a tool for describing, organizing, and exploring different types of expectations in ways that are useful in designing robots.

Some expectations were difficult to classify, particularly at higher levels of abstraction; for example believing that a robot is 'brave' does not neatly fit into a particular capability domain, though it can be weakly related to all three (physical, social, and computational). Many expectations spanned categories within the taxonomy, for example, expecting that a robot 'wants to shake hands' involves both physical and social domains. These challenges were perhaps a direct result of our aim for full coverage, resulting in encompassing definitions with some overlap, thus indicating potential for improved, more focused categories. In our current work, we represented this in our visualizations using increasingly blurred boundaries between the domains as the level of abstraction rises.

While our taxonomy provides full scope, we found limitations with expectation granularity. For example, expecting that a robot can walk across a room, or can see an item on a shelf, both fall into the *physical* and *operational* taxonomy bin despite being completely different expectations. Thus while our taxonomy provides a general lens for considering the range of possibilities, expectations specific to individual robots and interaction need more nuanced consideration. Within our framework, we intend for this to be addressed through the more intricate treatment offered by our cognitive process model.

Finally, while our framework focuses on classifying and explaining expectations, it does not directly address identifying which robot features may lead to what outcomes (e.g., if you add hands, how will people respond?). Thus important future work is to develop toolkits of robot features and designs that can be mapped (e.g., using experimental results) to desired expectation outcomes.

6.2.2 Theoretical Foundations

The rigour and reliability of our work is primarily rooted in the robustness of the theoretical works upon which it is constructed. Thus, the strength of the theoretical assumptions found in these works, and the empirical evidence that supports those assumptions, serve to bolster our own framework. Nonetheless, our contribution is merely an initial framework within a growing area of study, and moving forward we will need to continue to assess the theoretical assumptions to provide opportunities for future development.

A founding assumption of our framework is that people will respond to robots as if they were in some sense alive (*animorphism*, Section 3). Although a well-established stance in human-robot interaction, the reality is that robots are not alive, and we can reasonably expect limitations to the animorphism [116] that impact how people interact with robots. We found this in our expectation walkthrough example, where the fictional user was assumed to treat the robot in rigidly human terms (e.g., ‘friendly but [having] poor social etiquette’). In reality, a person may treat a robot as a new ontological category [116], applying animorphism while still treating it as a machine, which would impact application of our cognitive process. It remains to future work to better understand where this line lays.

One of the major theoretical anchors of our work is simulation theory. This is highlighted in our process mode (Figure 6) by the fact that all signals from the robot pass through the 'simulation as self' step. While considerable evidence exists to support this theory, *simulation theory* and its rival *theory theory* [sic] remain debated in human psychology [68], and thus this uncertainty extends to this part of our framework. Nonetheless, research supports the notion that cognitive simulations are at least a component of expectation formation [68].

Finally, we developed our taxonomy through qualitative analysis of a curated corpus. This could be strengthened through empirical work that specifically investigates how people understand expectations in relation to our categories. Further, our proposed inspection methods may be applied to a formally-developed corpus of user-elicited expectations to evaluate its analytical power in empirical applications.

6.2.3 Passive Role of the Human

Our cognitive process model is predominantly centered on how a person receives signals from a robot and the resulting internal cognitive processes. This treats the person primarily as a passive participant, as in our cognitive expectation walkthrough where the fictional user, confused by the robot's actions, struggles through interaction focused on internal cognitive processes. However, we may expect a person to additionally seek knowledge or prod the robot to explore abilities. Future work should more closely consider a person's active involvement in resolving expectation discrepancies.

6.2.4 Utility to Designers

We have demonstrated through case studies how our framework may be employed to support designers in managing user expectations of their robots (Section 5), as the only broad framework to date for expectations of robots. This framework provides new vocabulary, framing, and the first exploratory tools for designers to use to support exploration and understanding of robot expectations. Looking forward, we will need to work together with designers to study how they may employ our framework in practice, and identify opportunities to expand and refine the application of our framework through our proposed inspection methods. This will provide a more nuanced understanding of where and when our framework can support the process of robot design and evaluation.

7 Conclusion

Managing expectation discrepancy—where a person forms an inaccurate expectation of a robot, potentially leading to disappointment and interaction challenges—remains an open problem in human-robot interaction. In this paper we presented a comprehensive framework of expectation discrepancies, including a two-dimensional taxonomy for classifying expectations and a cognitive process of expectations that describes how people may form expectations of robots. We further developed and presented two novel analytical inspection methods for applying our framework in practice and exploring expectations in human-robot interaction. We use these inspection methods to demonstrate possible avenues by which the framework we developed can support designers to compare expectations across different robot designs, highlight areas of expectation discrepancy which may hinder interaction, and analyze and explain how those expectations emerge, although studying how designers may employ our framework in practice remains an important step for future work. In sum, our work provides some of the first frameworks and concrete tools for supporting robot creators in making informed choices to influence users' expectations of their robots.

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

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

The authors report there are no competing interests to declare.

References

[1] Breazeal C. Toward sociable robots. Robotics and Autonomous Systems. 2003;42(3–4):167–175.

[2] Schramm LT, Dufault D, Young JE. Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design. ACM/IEEE International Conference on Human-Robot Interaction. 2020;439–441.

[3] Komatsu T, Kurosawa R, Yamada S. How Does the Difference Between Users' Expectations and Perceptions About a Robotic Agent Affect Their Behavior? Int J of Soc Robotics. 2012;4(2):109–116.

[4] Sharkey A, Sharkey N. We need to talk about deception in social robotics! Ethics Inf Technol. 2021;23(3):309–316.

[5] Bruckenberger U, Weiss A, Mirnig N, et al. The Good, The Bad, The Weird: Audience Evaluation of a "Real" Robot in Relation to Science Fiction and Mass Media. In: Herrmann G, Pearson MJ, Lenz A, et al., editors. Social Robotics. Cham: Springer International Publishing; 2013. p. 301–310.

[6] Sandoval EB, Mubin O, Obaid M. Human Robot Interaction and Fiction: A Contradiction. In: Beetz M, Johnston B, Williams M-A, editors. Social Robotics. Cham: Springer International Publishing; 2014. p. 54–63.

[7] Kwon M, Jung MF, Knepper RA. Human expectations of social robots. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2016. p. 463–464.

[8] Natarajan M, Gombolay M. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2022 Aug 31]. p. 33–42. Available from: http://doi.org/10.1145/3319502.3374839.

[9] Cross ES, Ramsey R. Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions. Trends in Cognitive Sciences. 2021;25(3):200–212.

[10] Riek LD, Rabinowitch T-C, Chakrabarti B, et al. How anthropomorphism affects empathy toward robots. Proceedings of the 4th ACM/IEEE international conference on Human robot interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2009 [cited 2022 Aug 31]. p. 245–246. Available from: http://doi.org/10.1145/1514095.1514158.

[11] Kwon M, Huang SH, Dragan AD. Expressing Robot Incapability. Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. Chicago IL USA: ACM; 2018 [cited 2022 Aug 18]. p. 87–95. Available from: https://dl.acm.org/doi/10.1145/3171221.3171276.

[12] Paepcke S, Takayama L. 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. Osaka, Japan: IEEE Press; 2010. p. 45–52.

[13] Young JE, Hawkins R, Sharlin E, et al. Toward acceptable domestic robots: Applying insights from social psychology. International Journal of Social Robotics. 2009;1(1):95–108.

[14] Berzuk JM, Young JE. More Than Words: A Framework for Describing Human-Robot Dialog Designs. Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. IEEE Press; 2022. p. 393–401. Available from: https://doi.org/10.1109/HRI53351.2022.9889423.

[15] Mizumaru K, Satake S, Kanda T, et al. Stop Doing it! Approaching Strategy for a Robot to Admonish Pedestrians. ACM/IEEE International Conference on Human-Robot Interaction. IEEE Press; 2019. p. 449–457.

[16] Fujioka Y, Liu Y, Kanda T. I Need to Pass Through! Understandable Robot Behavior for Passing Interaction in Narrow Environment. Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2024 [cited 2024 Nov 6]. p. 213–221. Available from: https://dl.acm.org/doi/10.1145/3610977.3634951.

[17] Gervits F, Thielstrom R, Roque A, et al. It's About Time: Turn-Entry Timing For Situated Human-Robot Dialogue. Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue [Internet]. 2020 [cited 2021 Sep 29]. p. 86–96. Available from: https://aclanthology.org/2020.sigdial-1.12.

[18] Rosén J, Lindblom J, Billing E. The Social Robot Expectation Gap Evaluation Framework. In: Kurosu M, editor. Human-Computer Interaction Technological Innovation. Cham: Springer International Publishing; 2022. p. 590–610.

[19] Kahn PH, Freier NG, Kanda T, et al. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2008. p. 97–104. Available from: https://doi-org.uml.idm.oclc.org/10.1145/1349822.1349836.

[20] Bartneck C, Forlizzi J. A design-centred framework for social human-robot interaction. RO-MAN 2004 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No04TH8759) [Internet]. IEEE; 2004 [cited 2021 Sep 14]. p. 591–594. Available from: http://ieeexplore.ieee.org/document/1374827/.

[21] Yanco HA, Drury J. Classifying human-robot interaction: an updated taxonomy. 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat No04CH37583) [Internet]. IEEE; 2004 [cited 2021 Sep 9]. p. 2841–2846. Available from: http://ieeexplore.ieee.org/document/1400763/.

[22] Glas DF, Kanda T, Ishiguro H. Human-robot interaction design using interaction composer: Eight years of lessons learned. ACM/IEEE International Conference on Human-Robot Interaction. IEEE Press; 2016. p. 303–310.

[23] Reig S, Forlizzi J, Steinfeld A. Leveraging Robot Embodiment to Facilitate Trust and Smoothness. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2019. p. 742–744.

[24] Schaefer KE, Sanders TL, Yordon RE, et al. Classification of Robot Form: Factors Predicting Perceived Trustworthiness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2012;56(1):1548–1552.

[25] Rosenthal-von der Pütten AM, Krämer NC. How design characteristics of robots determine evaluation and uncanny valley related responses. Computers in Human Behavior. 2014;36:422–439.

[26] Bartneck C, Kulić D, Croft E, et al. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics. 2009;1(1):71–81.

[27] Carpinella CM, Wyman AB, Perez MA, et al. The Robotic Social Attributes Scale (RoSAS): Development and Validation. Proceedings of the 2017

ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2017 [cited 2022 Aug 31]. p. 254–262. Available from: http://doi.org/10.1145/2909824.3020208.

[28] Nomura T, Kanda T, Suzuki T, et al. Prediction of Human Behavior in Human–Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots. IEEE Transactions on Robotics. 2008;24(2):442–451.

[29] Phillips E, Zhao X, Ullman D, et al. 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). 2018. p. 105–113.

[30] Fortunati L, Manganelli AM, Höflich J, et al. Exploring the Perceptions of Cognitive and Affective Capabilities of Four, Real, Physical Robots with a Decreasing Degree of Morphological Human Likeness. Int J of Soc Robotics. 2023;15(3):547–561.

[31] Haring K, Matsumoto Y, Watanabe K. How Do People Perceive and Trust a Lifelike Robot? Proceedings of the World Congress on Engineering and Computer Science (WCECS). San Francisco, USA: IAENG; 2013. p. 23–25.

[32] Robinson FA, Velonaki M, Bown O. Smooth Operator: Tuning Robot Perception Through Artificial Movement Sound. Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2021 [cited 2022 Aug 31]. p. 53–62. Available from: http://doi.org/10.1145/3434073.3444658.

[33] van Maris A, Lehmann H, Natale L, et al. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2017 [cited 2022 Aug 31]. p. 313–314. Available from: http://doi.org/10.1145/3029798.3038435.

[34] Gompei T, Umemuro H. 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). 2015. p. 331–336.

[35] Hamacher A, Bianchi-Berthouze N, Pipe AG, et al. 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). 2016. p. 493–500.

[36] Mirnig N, Stollnberger G, Miksch M, et al. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. Frontiers in Robotics and AI [Internet]. 2017 [cited 2022 Aug 31];4.

[37] Ragni M, Rudenko A, Kuhnert B, et al. Errare humanum est: Erroneous robots in human-robot interaction. 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2016. p. 501–506.

[38] Salem M, Eyssel F, Rohlfing K, et al. To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. Int J of Soc Robotics. 2013;5(3):313–323.

[39] Xu K, Chen M, You L. The Hitchhiker's Guide to a Credible and Socially Present Robot: Two Meta-Analyses of the Power of Social Cues in Human–Robot Interaction. Int J of Soc Robotics. 2023;15(2):269–295.

[40] Eyssel F, Hegel F, Horstmann G, et al. Anthropomorphic inferences from emotional nonverbal cues: A case study. 19th International Symposium in Robot and Human Interactive Communication. 2010. p. 646–651.

[41] Stanton CJ, Stevens CJ. Don't Stare at Me: The Impact of a Humanoid Robot's Gaze upon Trust During a Cooperative Human–Robot Visual Task. Int J of Soc Robotics. 2017;9(5):745–753.

[42] Bryant D, Borenstein J, Howard A. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2022 Aug 31]. p. 13–21. Available from: http://doi.org/10.1145/3319502.3374778.

[43] Dennler N, Ruan C, Hadiwijoyo J, et al. Design Metaphors for Understanding User Expectations of Socially Interactive Robot Embodiments. J Hum-Robot Interact. 2023;12(2):21:1-21:41.

[44] de Graaf MMA, Ben Allouch S, van Dijk JAGM. What Makes Robots Social?: A User's Perspective on Characteristics for Social Human-Robot Interaction. In: Tapus A, André E, Martin J-C, et al., editors. Social Robotics. Cham: Springer International Publishing; 2015. p. 184–193.

[45] Lohse M. Bridging the gap between users' expectations and system evaluations. 2011 RO-MAN [Internet]. 2011 [cited 2024 Jan 22]. p. 485–490. Available from: https://ieeexplore.ieee.org/abstract/document/6005252.

[46] Salem M, Lakatos G, Amirabdollahian F, et al. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2015. p. 141–148. Available from: https://doi-org.uml.idm.oclc.org/10.1145/2696454.2696497.

[47] Collins EC, Prescott TJ, Mitchinson B, et al. MIRO: a versatile biomimetic edutainment robot. Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology [Internet]. New York, NY, USA: Association for Computing Machinery; 2015 [cited 2023 May 20]. p. 1–4. Available from: https://dl.acm.org/doi/10.1145/2832932.2832978.

[48] Goetz J, Kiesler S, Powers A. 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. 2003. p. 55–60.

[49] Olson JM, Roese NJ, Zanna MP. Expectancies. Social psychology: Handbook of basic principles. New York, NY, US: The Guilford Press; 1996. p. 211–238.

[50] Złotowski J, Sumioka H, Eyssel F, et al. Model of Dual Anthropomorphism: The Relationship Between the Media Equation Effect and Implicit Anthropomorphism. Int J of Soc Robotics. 2018;10(5):701–714.

[51] Epley N, Waytz A, Akalis S, et al. When We Need A Human: Motivational Determinants of Anthropomorphism. Social Cognition. 2008;26(2):143–155.

[52] Löffler D, Dörrenbächer J, Hassenzahl M. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2024 Jan 22]. p. 261–270. Available from: https://dl.acm.org/doi/10.1145/3319502.3374788.

[53] Burgess AM, Graves LM, Frost RO. My possessions need me: Anthropomorphism and hoarding. Scandinavian Journal of Psychology. 2018;59(3):340–348.

[54] Heider F, Simmel M. An Experimental Study of Apparent Behavior. The American Journal of Psychology. 1944;57(2):243–259.

[55] Nass C, Moon Y. Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues. 2000;56(1):81–103.

[56] Seo SH, Geiskkovitch D, Nakane M, et al. 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). 2015. p. 125–132.

[57] Sanoubari E, Seo SH, Garcha D, et al. 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) [Internet]. IEEE; 2019 [cited 2022 May 11]. p. 382–391. Available from: https://ieeexplore.ieee.org/document/8673326/.

[58] Seo SH, Griffin K, Young JE, et al. Investigating People's Rapport Building and Hindering Behaviors When Working with a Collaborative Robot. International Journal of Social Robotics. 2018;10(1):147–161.

[59] Meltzoff AN, Brooks R, Shon AP, et al. "Social" robots are psychological agents for infants: A test of gaze following. Neural Networks. 2010;23(8):966–972.

[60] Paré G, Kitsiou S. Chapter 9 Methods for Literature Reviews. Handbook of eHealth Evaluation: An Evidence-based Approach [Internet] [Internet]. University of Victoria; 2017 [cited 2025 May 22]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK481583/.

[61] Stefanidi E, Bentvelzen M, Woźniak PW, et al. Literature Reviews in HCI: A Review of Reviews. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2023 [cited 2025 May 22]. p. 1–24. Available from: https://dl.acm.org/doi/10.1145/3544548.3581332.

[62] Holthaus P, Schulz T, Lakatos G, et al. Communicative Robot Signals: Presenting a New Typology for Human-Robot Interaction. Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: Association for Computing Machinery; 2023 [cited 2023 Mar 16]. p. 132–141. Available from: https://dl.acm.org/doi/10.1145/3568162.3578631.

[63] Hall S, Hobson D, Lowe A, et al. Encoding/decoding. Culture, Media, Language [Internet]. London, UNITED KINGDOM: Taylor & Francis Group; 1980 [cited 2023 Feb 9]. p. 117–127. Available from: http://ebookcentral.proquest.com/lib/umanitoba/detail.action?docID=179321.

[64] Wodehouse A, Brisco R, Broussard E, et al. Pareidolia: characterising facial anthropomorphism and its implications for product design. Journal of Design Research. 2018;16(2):83–98.

[65] Burgoon JK, Jones SB. Toward a Theory of Personal Space Expectations and Their Violations. Human Communication Research. 1976;2(2):131–146.

[66] Burgoon JK. Expectancy Violations Theory. The International Encyclopedia of Interpersonal Communication [Internet]. John Wiley & Sons, Ltd; 2015 [cited 2023 Feb 9]. p. 1–9. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118540190.wbeic102.

[67] Burgoon JK, Hale JL. Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. Communication Monographs. 1988;55(1):58–79.

[68] Shanton K, Goldman A. Simulation theory. WIREs Cognitive Science. 2010;1(4):527–538.

[69] Gordon RM. Folk Psychology as Simulation. Mind & Language. 1986;1(2):158–171.

[70] Gallese V, Goldman A. Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Sciences. 1998;2(12):493–501.

[71] Gordon RM. The simulation theory: Objections and misconceptions. Mind & Language. 1992;7(1–2):11–34.

[72] Tamir DI, Mitchell JP. Anchoring and adjustment during social inferences. Journal of Experimental Psychology: General. 2013;142(1):151–162.

[73] Epley N, Keysar B, Van Boven L, et al. Perspective Taking as Egocentric Anchoring and Adjustment. Journal of Personality and Social Psychology. 2004;87(3):327–339.

[74] Ross L, Ward A. Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. Values and Knowledge. Psychology Press; 1996.

[75] Mitchell P, Robinson EJ, Isaacs JE, et al. Contamination in reasoning about false belief: an instance of realist bias in adults but not children. Cognition. 1996;59(1):1–21.

[76] Birch SAJ, Bloom P. The Curse of Knowledge in Reasoning About False Beliefs. Psychol Sci. 2007;18(5):382–386.

[77] Ames DR. Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. Journal of Personality and Social Psychology. 2004;87(3):340–353.

[78] Krueger JI. From social projection to social behaviour. European Review of Social Psychology. 2007;18(1):1–35.

[79] Meltzoff AN. 'Like me': a foundation for social cognition. Developmental Science. 2007;10(1):126–134.

[80] Epley N, Waytz A, Cacioppo JT. On seeing human: A three-factor theory of anthropomorphism. Psychological Review. 2007;114:864–886.

[81] Dourish P. Where the Action is: The Foundations of Embodied Interaction. MIT Press; 2001.

[82] Streeck J, Goodwin C, LeBaron C, editors. Embodied Interaction: Language and Body in the Material World. Cambridge University Press; 2011.

[83] Ziemke T. What's that thing called embodiment? Proceedings of the Annual Meeting of the Cognitive Science Society. 2003.

[84] Hoggenmueller M, Chen J, Hespanhol L. Emotional expressions of non-humanoid urban robots: the role of contextual aspects on interpretations. Proceedings of the 9TH ACM International Symposium on Pervasive Displays [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2022 Aug 31]. p. 87–95. Available from: http://doi.org/10.1145/3393712.3395341.

[85] Carter MJ, Fuller C. Symbolic interactionism. Sociopedia [Internet]. 2015 [cited 2023 Mar 2]; doi: 10.1177/205684601561.

[86] Young JE, Sung J, Voida A, et al. Evaluating Human-Robot Interaction: Focusing on the Holistic Interaction Experience. International Journal of Social Robotics. 2011;3(1):53–67.

[87] Thiessen R, Rea DJ, Garcha DS, et al. 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). 2019. p. 11–18.

[88] Gazzola V, Rizzolatti G, Wicker B, et al. The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. NeuroImage. 2007;35(4):1674–1684.

[89] Hoenen M, Lübke KT, Pause BM. Non-anthropomorphic robots as social entities on a neurophysiological level. Computers in Human Behavior. 2016;57:182–186.

[90] Oberman LM, McCleery JP, Ramachandran VS, et al. 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. 2007;70(13):2194–2203.

[91] Hannibal G. The Trust-Vulnerability Relation - A Theory-driven and Multidisciplinary Approach to the Study of Interpersonal Trust in Human-Robot Interaction [Internet] [Thesis]. Technische Universität Wien; 2023 [cited 2023 Jun 1]. Available from: https://repositum.tuwien.at/handle/20.500.12708/139926.

[92] Richardson K. An Anthropology of Robots and AI: Annihilation Anxiety and Machines. New York: Routledge; 2015.

[93] Lemaignan S, Fink J, Dillenbourg P, et al. The Cognitive Correlates of Anthropomorphism. HRI 2014 Workshop: HRI: a bridge between Robotics and Neuroscience [Internet]. 2014. Available from: http://infoscience.epfl.ch/record/196441.

[94] Xu J, Howard A. 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). 2018. p. 435–441.

[95] Paetzel M, Perugia G, Castellano G. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2023 Feb 9]. p. 73–82. Available from: https://doi.org/10.1145/3319502.3374786.

[96] Aronson E, Willerman B, Floyd J. The effect of a pratfall on increasing interpersonal attractiveness. Psychon Sci. 1966;4(6):227–228.

[97] Mulrow CD. Systematic Reviews: Rationale for systematic reviews. BMJ. 1994;309(6954):597–599.

[98] Buxton B. Sketching User Experiences: Getting the Design Right and the Right Design. Amsterdam Boston: Morgan Kaufmann; 2007.

[99] Braun V, Clarke V. Thematic Analysis: A Practical Guide. SAGE Publications; 2021.

[100] Harboe G, Huang EM. Real-World Affinity Diagramming Practices: Bridging the Paper-Digital Gap. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2015 [cited 2024 May 13]. p. 95–104. Available from: https://dl.acm.org/doi/10.1145/2702123.2702561.

[101] O'Connor C, Joffe H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. International Journal of Qualitative Methods. 2020;19:1609406919899220.

[102] Aldebaran. Pepper the humanoid and programmable robot |Aldebaran [Internet]. [cited 2023 Sep 29]. Available from: https://www.aldebaran.com/en/pepper.

[103] aibo [Internet]. [cited 2022 Aug 15]. Available from: https://us.aibo.com/.

[104] Moon Aj, Calisgan E, Bassani C, et al. The Open Roboethics initiative and the elevator-riding robot. In: Calo R, Froomkin AM, Kerr I, editors. Robot Law [Internet]. Cheltenham, UK: Edward Elgar Publishing; 2016. p. 131–162. Available from: https://www.elgaronline.com/view/edcoll/9781783476725/9781783476725.00014.xml.

[105] Cha E, Dragan AD, Srinivasa SS. Perceived robot capability. 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2015. p. 541–548.

[106] Kuzminykh A, Sun J, Govindaraju N, et al. Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2022 Aug 31]. p. 1–13. Available from: http://doi.org/10.1145/3313831.3376665.

[107] Rogers Y, Sharp H, Preece J. Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons; 2023.

[108] Ledo D, Houben S, Vermeulen J, et al. Evaluation Strategies for HCI Toolkit Research. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2018 [cited 2025 May 20]. p. 1–17. Available from: https://dl.acm.org/doi/10.1145/3173574.3173610.

[109] Atlas [Internet]. Boston Dynamics. [cited 2025 Oct 2]. Available from: https://bostondynamics.com/atlas/.

[110] Passler Bates D, Young JE. SnuggleBot: A Novel Cuddly Companion Robot Design. Proceedings of the 8th International Conference on Human-Agent Interaction [Internet]. New York, NY, USA: ACM; 2020 [cited 2022 May 11]. p. 260–262. Available from: https://dl.acm.org/doi/10.1145/3406499.3418772.

[111] PARO Therapeutic Robot [Internet]. [cited 2022 Aug 15]. Available from: http://www.parorobots.com/.

[112] GROOVE X, Inc. LOVOT [Internet]. [cited 2025 May 16]. Available from: https://lovot.life/en/.

[113] Rogers Y, Sharp H, Preece J. Chapter 15. Analytical Evaluation. Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons; 2023.

[114] Hevner AR, March ST, Park J, et al. Design Science in Information Systems Research. MIS Quarterly. 2004;28(1):75–105.

[115] Passler Bates D, Dudek SY, Berzuk JM, et al. SnuggleBot the Companion: Exploring In-Home Robot Interaction Strategies to Support Coping With Loneliness. Proceedings of the 2024 ACM Designing Interactive Systems Conference [Internet]. New York, NY, USA: Association for Computing Machinery; 2024 [cited 2024 Jul 2]. p. 2972–2986. Available from: https://doi.org/10.1145/3643834.3660702.

[116] Kahn PH, Reichert AL, Gary HE, et al. The new ontological category hypothesis in human-robot interaction. Proceedings of the 6th international conference on Human-robot interaction - HRI ’11 [Internet]. Lausanne, Switzerland: ACM Press; 2011 [cited 2022 Aug 3]. p. 159. Available from: http://portal.acm.org/citation.cfm?doid=1957656.1957710.

Appendix A. Systematic Expectation Dissection Case Studies

We demonstrate systematic expectation dissection with four real example robots as case studies: the SoftBank Pepper [102], Boston Dynamics' Atlas [109], the Sony aibo [103], and SnuggleBot [110]. To help demonstrate the technique, we informally generated example expectations that a hypothetical user may have of these robots (Table 2).

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

Table 2. A dummy 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 labelled plot symbols in the example visualizations (Figures 12-15).

Our first example robot is the SoftBank Pepper [102]. 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. We took our hypothetical user expectations (Table 2) and plotted them onto our taxonomy space (Figure 12); the visual overview provides quick insight into common expectation patterns in the form of clusters of plot points, as well as conspicuously empty regions. 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 hypothetical 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.

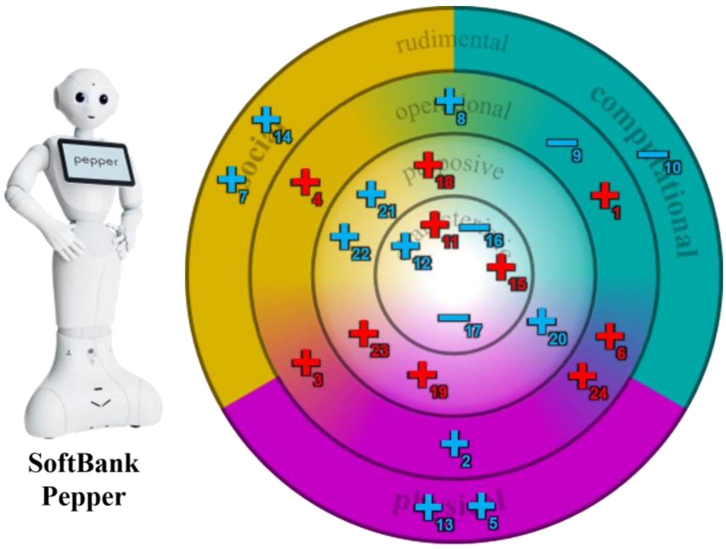


Figure 12. Example expectations of the SoftBank Pepper [102] visualized with our expectations taxonomy.

Our next example robot is Boston Dynamics' Atlas [109], a humanoid robot which walks on two legs and is capable of picking up and manipulating objects. The Atlas, as another humanoid robot, is of a similar form factor to Pepper, but with key design differences such as its legs and its lack of a human-like face. The visualization (Figure 13) highlights that, in comparison with Pepper, the user had more expansive physical expectations of Atlas, which were largely matched with its capabilities, but had similar social expectations (which were largely mismatched). This may suggest that the unique aspects of Atlas' design were effective at increasing expectations (perhaps its legs encouraged greater physical expectations), but not as effective at encouraging more realistic social expectations (its lack of 'face' did not discourage the user from thinking it could speak). In this way, comparing the visualizations generated through systematic expectation dissection of similar designs may highlight the impacts of the smaller design differences on user expectations.

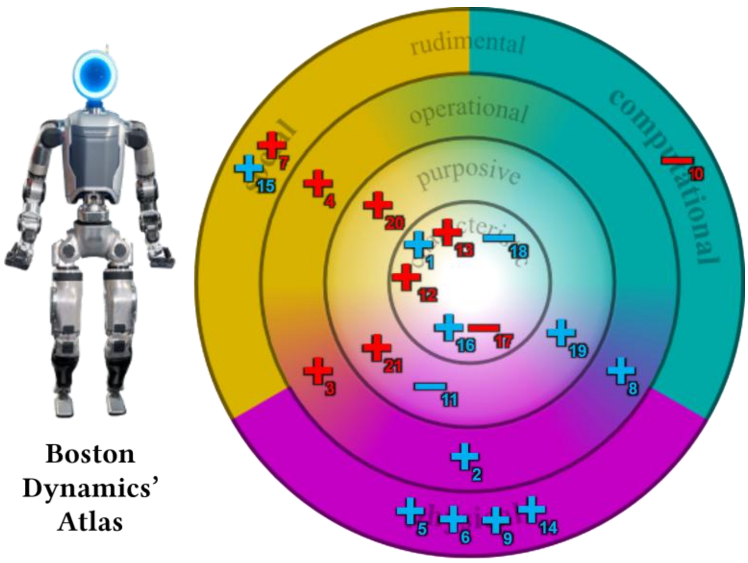


Figure 13. Example expectations of the Atlas [109] visualized with our expectations taxonomy.

Our next example robot is the Sony aibo robotic dog designed to fulfill the role of a pet in a user's home [103]. We again plotted the expectations in Table 2 onto our taxonomy space (Figure 14). 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).

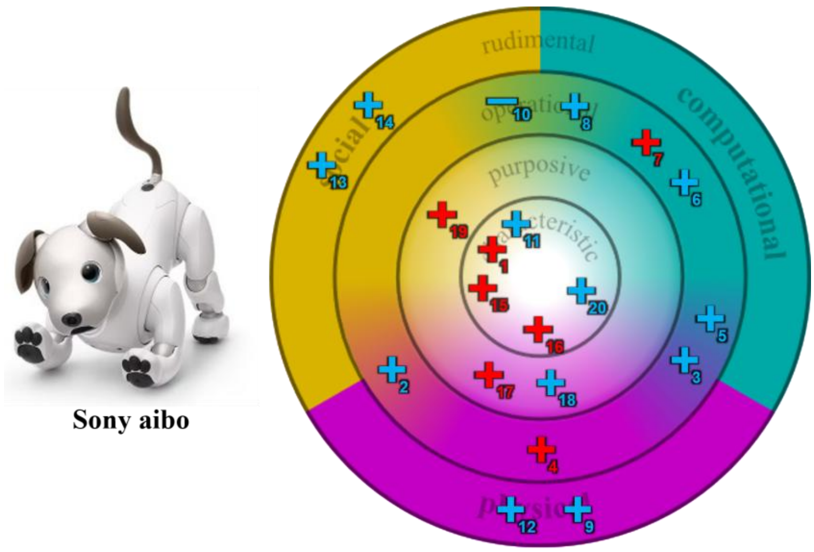


Figure 14. Example expectations of the Sony aibo [103] visualized with our expectations taxonomy.

Our final example robot is SnuggleBot [110] (Figure 15), a stuffed narwhal with lights, mobile limbs, and sensors, which is designed to provide companionship to users [110]. One immediate difference with this visualization is that, compared to the other three 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).

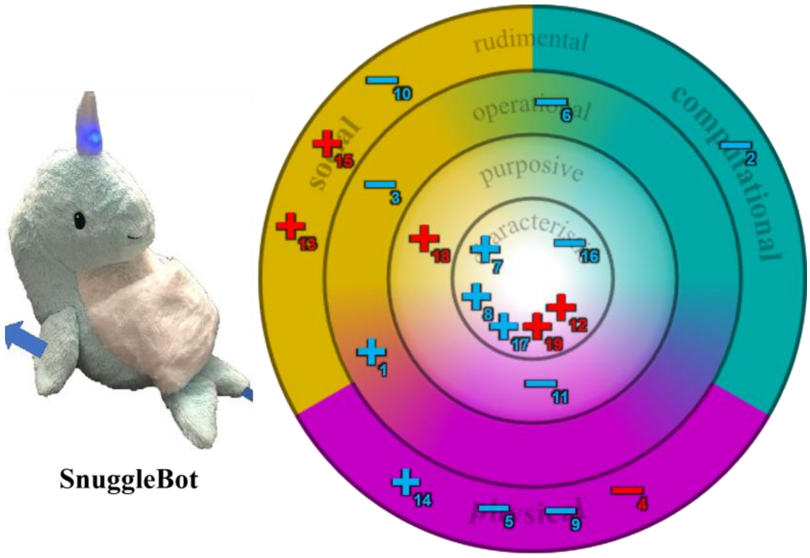


Figure 15. Example expectations of the SnuggleBot [110] visualized with our expectations taxonomy.

Appendix B. Cognitive Expectation Walkthrough Case Study

Robot — For this demonstration we continue with the SoftBank Pepper 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 7). There is a small sign next to the tablet instructing people to touch it to start.

Walkthrough

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).

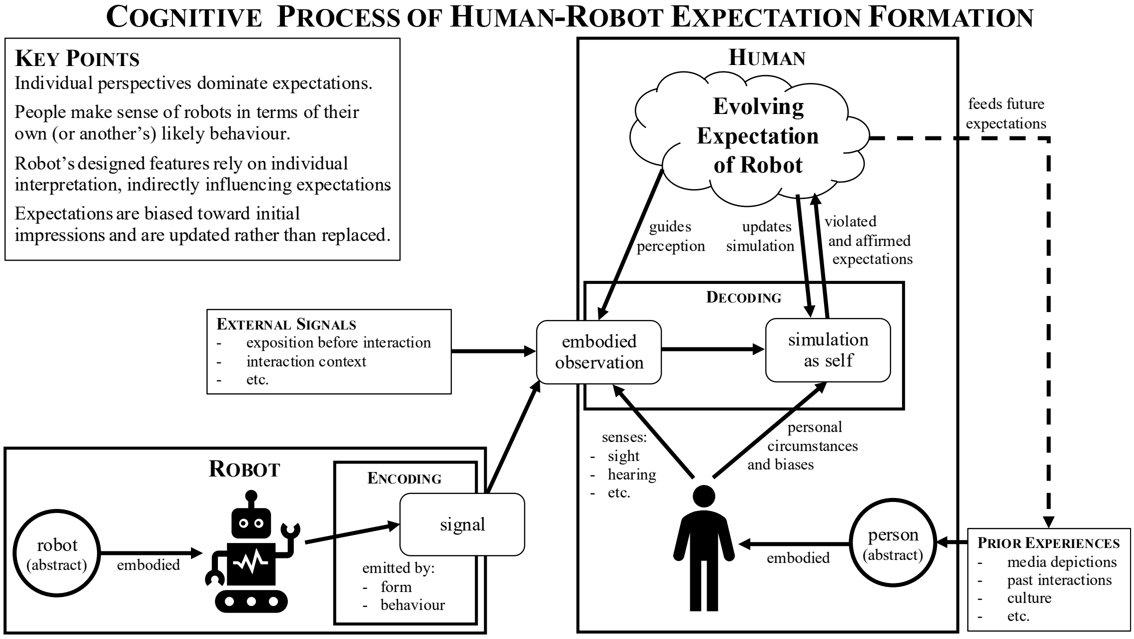
From an embodied observation point of view, Sam notices the visuals more than the audio given the noisy scenario. Sam applies her existing experience of seeing the robot on the news to interpret these signals, and combined with her existing expectations of robots (animorphic) she did not notice the tablet 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.

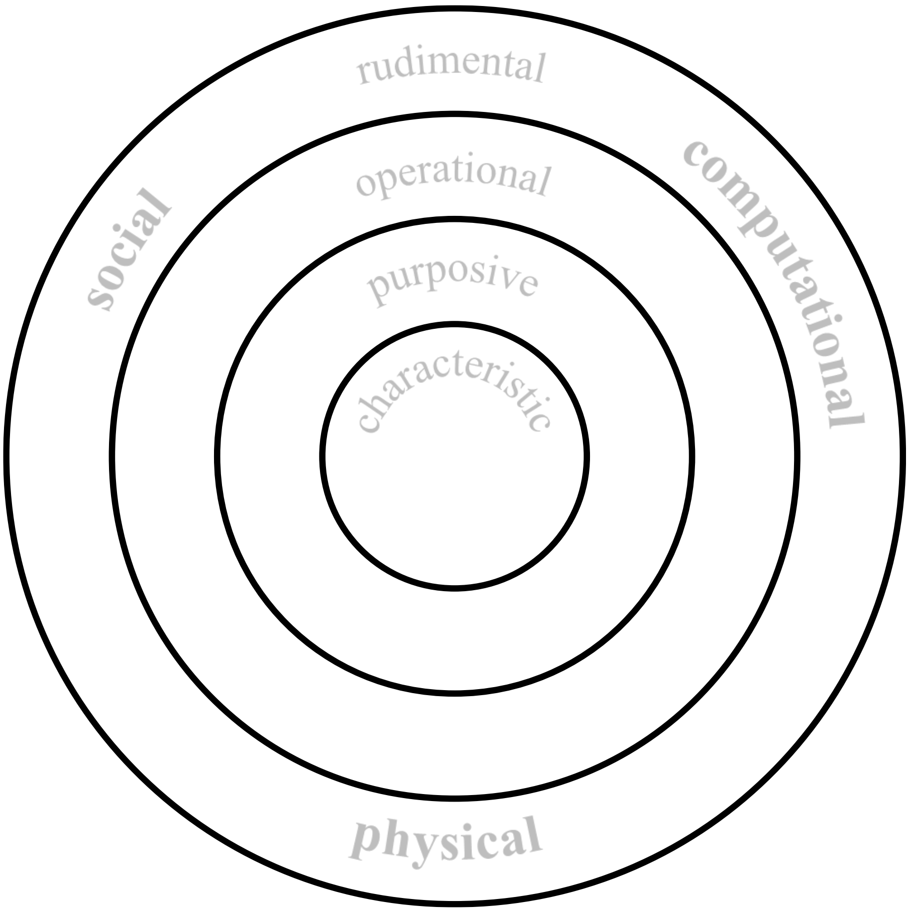
Sam approaches the robot and waves, saying hello. The robot does not respond. Observing this response signal with her existing expectations, Sam is surprised. Simulating this reaction, Sam initially wonders if the robot is simply unfriendly, violating her expectations, but then realizes the robot maybe 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 expectation 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.

Appendix C. Systematic Expectation Dissection Printout





Expectations

1. ____________________
2. ____________________
3. ____________________
4. ____________________
5. ____________________
6. ____________________
7. ____________________
8. ____________________
9. ____________________
10. ____________________
11. ____________________
12. ____________________
13. ____________________
14. ____________________
15. ____________________
16. ____________________
17. ____________________
18. ____________________
19. ____________________
20. ____________________
21. ____________________
22. ____________________
23. ____________________
24. ____________________
25. ____________________