UNIVERSITY OF CALIFORNIA
SANTA CRUZ

THE CASE OF THE CURIOUS ROBOT: ON THE VIABILITY OF CURIOUS BEHAVIOR IN ROBOTS

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTATIONAL MEDIA

by

Kevin Weatherwax

September 2021

The Thesis of Kevin Weatherwax
is approved:

Professor Leila Takayama, Chair

Professor Steve Whittaker

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

The Case of the Curious Robot: On the Viability of Curious Behavior in Robots

by

Kevin Weatherwax

Curiosity is a core drive for learning in humans which is increasingly being looked at in developing robots capable of internally motivated lifelong learning. However, there are also potential risks associated with robots being curious when people do not expect or want robots to be curious, especially when the robot already has work tasks to perform. Further, how robots' mental states are described may prime expectations in human counterparts. This thesis presents a pair of experiments on people's perceptions of four levels of curious robot behavior, designed by a professional animator. We also explored the impacts of a curious robot being seen as on-duty vs. off-duty. In addition, we examined whether our curious robot behavior matched peoples' expectations when primed to expect a "curious" vs. "learning" vs. "autonomous" robot. In both studies, as curious behavior increased so to did ratings of the robots' thinking ability coupled with a decrease in ratings of it as an effective working and social agent, particularly when it was framed as on-duty. Further, each cognitive prime resulted in a unique trajectory of expectation matching showing that robots' internal mental states should be matched with analogous external behaviors. Our findings have implications for the design and development of robots that engage in learning modeled on human cognition.

Acknowledgments

As is true with all research efforts the work presented here is the outcome of a collaborative effort on several fronts and I would like to acknowledge the people who helped along the way. First and foremost, however, is Prof. Leila Takayama, who co-authored and supervised the work presented in this thesis. Also, our top notch research assistant Elizabeth Trimble, who assisted in managing online MTurk recruitment, qualitative data analysis, and figures. And perhaps most of all, Douglas Dooley the master animator, who provided tireless collaboration in addressing endless notes and input from this researcher.

I would also like to thank some dear friends and fellow graduate students who helped me carry on in times when I wanted to stop. Tessa “the fleece vest” Eagle who provided assistance as a data viz wiz to create lovely tables and charts. Her data visualization work can be found @tessuheagle. Also Lee ”Tabee” Taber who gave thoughtful edits to the final document and served as a reasonable sounding board to my Kaczynski-esque study designs for the past 18 months. More of his work can be found at leetaber.com.

I would also like to acknowledge Oakland based illustrator Louis Riddick who contributed his artistic talent in producing a highly accurate visualization of my thematic codes in Study 2. You can see more of his work on instagram @murkfiend and at louisriddick.com.

Additionally, I would like to thank the following people. Prof. Steve Whittaker, who provided guidance throughout the writing process and Prof. Douglas Bonnett, who lent his time and mastery of inferential statistics to me throughout the data analysis process.

This work was funded by Honda Research Institute, in collaboration with the University of Washington, as part of the Curious Minded Machine (CMM) initiative.

Chapter 1

Introduction

There is growing interest in creating robots that are capable of lifelong learning, enabling robots to expand beyond their preprogrammed capabilities over time [26,47,66]. However, traditional approaches to artificial learning for non-human agents, while impressive in their ability to find and reproduce abstract patterns in large data-sets, have struggled with operationalizing what humans consider to be actual learning. This is because human learning isn't simply a product of input and output but rather it is the outcome of motivation through a complex interplay of physiology, emotion, and cognition, all situated in–and acting in response to–the world [29, 45, 46, 58, 59]. At the center of that motivation, driving those other facets of learning, is curiosity. For many artificial intelligence (AI) researchers curiosity is seen as a possible missing link in developing true "thinking machines" (i.e., robot learning algorithms) that are inspired by various conceptions of human of curiosity that drive intrinsically-motivated robot learning over time [9, 12, 13, 43, 46, 67]. This is because curiosity, as a motivational manager of learning, can drive the discovery, understanding, mastery, and contextualization of "rare or deceptive rewards" [45]. Curious robots could conceivably seek out, identify, employ, and refine new knowledge and skills in real world settings. It is hoped, then, that developing robot curiosity could enable us to create more robust agents by imbuing them with the ability to adapt to changing environments as well as free robots from the tether of human expertise (e.g., operators, controllers, wranglers, etc.).

While curiosity is a potentially useful construct for developing machines capable of life-long learning, the implications for human-robot interactions (HRI) are less clear. There is a risk that these curious robots might not develop into systems that match the abilities [2] or values [11] of the people who live and work around them, especially if their responses to these robots are not taken into account. It may be especially important to reconcile internal functions of curiosity (e.g., subject of focus, interest, intent) with appropriate external representations of the same social and behavioral states, as demonstrated with expressive robots in the past [64]. Using models of human cognition to equip robots with adaptable learning abilities and enabling them to explore a wider range of environments raises new questions and challenges for HRI research and design. Just a few of these questions include:

- Will (untrained) people recognize curiosity in a robot?
- Is there such a thing as having too much robot curiosity?
- In which situations will a curious robot be acceptable vs. unacceptable to the human stakeholders around it?
- Does a robot's internal state (mind) need to match its external behaviors (body)?

Essentially, we must also understand the external aspects of curiosity in robots; how it is experienced, perceived, interpreted in robotic form(s) and how we design for it. While research in human-robot interaction (HRI)–and in psychology more generally–has explored some of these questions directly, and tangentially by studying human to human relationships, many important gaps remain. Without this attention to these broader HRI issues, curious robots could be poorly received, being seen as intrusive, creepy, or threatening, regardless of ability or functional usefulness [53]. It would be a shame if a curious robot failed only because it was unable to manage the social norms and expectations of a less "robotic" mind.

This thesis explores how curious robot behaviors, are perceived and interpreted by untrained end-users, who might one day encounter robots in their workplaces. Here we evaluate how people's responses to curious robot behaviors change with the context of the robot being busy performing a task for someone (on-duty) or exploring on its own time (off-duty). We measured people's perceptions of the robot, including its curiosity level, drive to learn, likability, self-sufficiency, focus, competence, and responsibleness. From an HRI design perspective, our studies show that robot curiosity can be communicated effectively through robot behaviors, using the actual degrees of freedom and joint limits of a real mobile robot. Moreover, our empirical findings regarding the boundaries of acceptable robot behaviors have implications (and raise more questions) for researchers developing robots that seek to learn.

Chapter 2

Of Curiosity, Humans, and Robots

To situate the current work, we present a brief overview of curiosity, then review where we stand in developing computational curiosity for learning agents, how curiosity is expressed by people, and how behavior design can be used to express robot curiosity. We explain how curiosity drives human learning and has become an area of focus for algorithmic approaches aimed at imitating the same kind of adaptive learning humans are capable of. We also discuss how curiosity, in human centered research, is not a wholly internal process but also includes external signals that trigger social interactions which facilitate the learning process. Computational approaches to instantiating curiosity in non-human agents (e.g., robots) has, conversely, not considered these important social and behavioral facets. Finally, we discuss how curious–and other learning based–robots could be designed more thoughtfully by aligning their internal (mind) and external (body) states.

2.1 What is curiosity?

Curiosity is a complex construct and a fundamental facet of what drives human learning. The prevalent view, informed by Loewenstein [37], conceptualizes curiosity as a kind of motivating drive for exploration and learning in humans and animals [15,31,32, 45]. In particular Loewenstein [37] framed the catalyst of curiosity as the realization of a "knowledge gap" which an agent then attempts to fill. To do this, individuals exploit previously obtained knowledge/skills to engage in a mode of learning that involves an active search for, and exploration of, novel situations and information. With this in mind we can frame curiosity as a cycle whereby an agent; (1) recognizes an **information gap**, (2) develops the **motivation** to fill it with relevant information, (3) actively **seeks knowledge** in an area that makes the most sense based on previous discoveries, and (4) **rewarded** by the acquisition of new knowledge that can be exploited, seeks new avenues for exploration.

However, under this overarching model there are competing explanations for what triggers curiosity (i.e., stimulation by environmental factors vs. abstract cognition) and for how curiosity motivates action (i.e., interest vs. deprivation). Different research domains that have adopted the topic of curiosity have developed new and sometimes domain-specific terms, or have renamed and re-purposed existing concepts. Curiosity research in human psychology is usually grounded in relation–or opposition–to Berlyne's two dimensional model of curiosity [32] represented on the axes of *Perceptual to Epistemic* [4] and *Specific to Diversive* [3]. In this framing, "perceptual curiosity" is rooted in environmental stimuli that register through an agent's senses. Novel things that are seen, heard, smelled, etc., draw the figurative eye, inciting interest and increasing familiarity with the source. "Epistemic curiosity" on the other hand is a state of the mind, the desire for knowledge and learning which can be either abstract or concrete. While the first of Berlyne's axes distinguishes the foci of curiosity, the second axis describes modes of curiosity. "Specific curiosity", as the name suggests, is interest in a single, specific piece of knowledge or sensation while "diversive curiosity" is not tied to anything in particular, and stems from a general desire to assuage a lack of stimulation [37]. For example, reading a book on a topic that has caught one's interest would be specific, while roaming around a library looking for an interesting book would be diversive curiosity.

Then there are two schools of explanation for what triggers curiosity; (1) drive theories and (2) optimal-level arousal theories [36]. These differ regarding whether it is best to describe/conceive the function of curiosity as either; (a) a reduction of deprivation or (b) as an induction of interest, respectively. Drive theories frame curiosity as being similar to other drives (e.g., hunger, thirst, fear). Thus, drive curiosity is conceived of as an unpleasant sensation brought on by a sense of uncertainty which then leads an agent to act (e.g., explore, learn, interact), thereby reducing this feeling with new information. For example, this could look like sitting bored on a couch channel surfing to find something interesting or googling a specific piece of information because not knowing is making you frustrated. Reward is accrued as increased exposure to the novel stimuli reduces the negative sensation of curiosity. Conversely, optimal-arousal theorists argue that the “curiosity as a drive” theory fails to account for agents who willingly seek out novel situations, thereby inciting, rather than avoiding, curiosity inducing situations. Thus, they propose that curiosity is a pleasurable sensation, which is why agents willingly induce it, and then attempt to maintain it at an optimal level (i.e., not bored but also not overwhelmed). More to the point though, is that these frameworks are the basis for most computational models of curiosity for non-human agents.

2.2 Computational Curiosity and Robots

As we described above, curiosity is an important building block for human learning [3, 4, 32, 36, 57], which makes it a promising area of development for machine learning. This is due, specifically, to curiosity’s ability to facilitate diverse and divergent learning over long periods of time and then assemble a tapestry of rich, complex, and flexible knowledge. Moreover, the resulting knowledge is, in theory, unique to the needs of individual robotic agents depending on its “lived” experiences and settings. With traditional reinforcement learning (RL), the learning agent is guided by reward functions associated with specific goals then develops, evaluates, and refines policies for optimal reward. Singh et al., [61] used a psychological and evolutionary lens to compare these common RL practices as an abstraction and operationalization of extrinsic motivation based learning found in nature (e.g., animals, humans). Herein, a learning agent wholly driven by extrinsic reward functions will pursue, learn, and perform, behaviors directly related to those ascribed goals but will, of course, not pursue or value experiences which don't pay those off. Essentially they argue that algorithms that include models of curiosity, will increase the overall lifelong "fitness" of a learning agent as it will develop a richer understanding of its environment(s), as well as novel abilities to impact its environment(s) [60, 61]. So a traditional RL based agent would need to be set some kind of predefined goal and then iterate its approach over time to improve performance but that experience cannot be applied to new or disparate tasks. A similar agent that included a model of curiosity based motivation could, theoretically, discover and master the same new task, without needing to be set to that task specifically, but could also discover and catalogue novel abilities that may be of value in, as of yet, unknown future challenges for it. We believe this framing helps us understand the value of curiosity based learning agents in the context of wanting to bring them out of the lab and in to the streets, so to speak.

For humans, curiosity is quite a recognizable trait. People can easily tell when someone is curious, what they are curious about, and–perhaps most importantly–whether they can/should intervene to provide assistance. But what does a curious robot look like? In a review of a curiosity based learning for robots, Oudeyer highlighted a demonstration of a curious robot arm which freely explored its environment by interacting with a pair of joysticks in front of it which controlled the movements of mechanical arm confined inside a small play environment with lights, sounds, and objects it could affect change on. This robot was able to discover novel effects it could have on the environment and then practice recreating these until the novelty transformed to mastery of new skills [45]. While this is a remarkable demonstration (See Oudeyer's demo: here) of the potential of intrinsic-motivated learning (i.e., curiosity), to an outside observer the robot arm seems to grope about the surrounding space for long periods of time in erratic jerking movements. This specific example raises an important problem central to this thesis. Namely, curiosity does not exist in a vacuum.

Human curiosity is, in large part, facilitated by interactions with other agents in the environment (e.g., teachers, peers, mentors), not simply in response to it. So while progress has been made on developing robots capable of curiosity-driven learning, the source of mentoring has largely been the scientists who develop it. They provide parameters, assistance, direction, and guidance as well as tightly control the environment to facilitate success, much like a preschool teacher, to their robots because they are just that; their robots. In real world settings, this relationship may fall apart when robots are not designed to take likely social and organizational contexts into account [41]. Gordon, a leading researcher in curiosity and robotics, offers a poignant assessment on the strengths and limitations of what robots can learn in non-social vs. social settings using curiosity based algorithms [13]. Namely, curious robot can learn very well in isolation but only about itself while social settings introduce far more complex and dynamic factors [13]. Gordon goes on to stress how the success of robot curiosity, in a social setting, is largely dependent on how people respond to the robot in its efforts to learn.

Moreover, the only prior work that has directly examined human perceptions of a curious robot (rather than implementations of curiosity based learning for robots) identified that curious deviations during a prescribed task resulted in negative considerations of the robot [69]. This is especially interesting considering that much of the ongoing research on developing curious robots assumes curiosity to be a universally positive trait for robots. These kinds of blanket assumptions of robot usefulness have proven to be detrimental in the past. For example, an ethnographic study of a delivery robot deployed in various hospital settings found that in a high impact trauma unit was seen as a burdensome nuisance but in a maternity ward was a welcome addition [41]. Furthermore, these hospital robots were sabotaged, abused, and stuffed into utility closets in the trauma center but received warmly and viewed as useful in the postpartum center [41]. We can infer then, that for robots to engage in lifelong learning in human-populated spaces (e.g., offices, hospitals, airports, shopping malls), it will be similarly important for us to figure out when, where, and how robots should engage in curiosity-driven exploration in a way that will be acceptable for human co-workers and bystanders. One potentially fruitful starting point for developing socially acceptable curious robot behaviors is to look for inspiration in how people express curiosity.

2.3 Expressing Human Curiosity

While there are many nuanced facets of human curiosity (e.g., Kashdan's five dimensions of curiosity [30]), exploratory and investigative behaviors are its most readily recognizable components. The internal cognitive processes of curiosity often manifest externally as focus and attention (e.g., interest, specific curiosity) or wandering and seeking new experiences (e.g., exploration, divisive curiosity) [3, 4]. It is important to note that research on curiosity is tightly intertwined with research on interest [54,57–59] so interest is included in some of the behaviors described below.

Looking to past work on curiosity is helpful for describing what it means for someone—or some robot—to behave in a curious manner. As such, we focus on what observable behaviors have been operationalized or measured in curiosity related research. One of the most clear cut curious behaviors is exploration. At its core, exploration refers to investigative behaviors (e.g., seeking, searching, manipulating) relating to environmental stimuli (e.g., specific things, exploring spaces) and is often the focus of child developmental work [6, 27, 49].

Facial expressions and head tilting are also useful signals for recognizing curiosity in others. In a pair of experiments, Reeve [51, 52] captured changes in facial expressions when people were shown interesting stimuli. A more recent empirical study, aimed at developing curiosity auto-recognition, examined videos of children primed to feel curious about objects (novel vs. mundane); when participants were primed to feel curious, the researchers noticed changes in the gaze, head shifts (i.e., panning and tilting), and moving around the object to focus on different parts of it [44].

Another strong behavioral signal of curiosity, in humans, is spending more time observing and becoming familiar with novel objects [44]. This is consistent with prior research on curious behavior in children, in which reacting to new stimuli by moving towards it and spending more time examining it was indicative of high curiosity [38–40].

As each robot has a different set of affordances, these expressions of human curiosity cannot be mapped directly onto any given robot, but they can be used to inspire the design of curious robot motions.

2.4 Expressing Robot Curiosity

Robot expressive motions are a fundamental aspect of human-robot interaction, especially because a robot's motions influence how people perceive and interact with it [16,21]. People read meaning into the motion of abstract shapes [17], not just anthropomorphic agents. As such, it is important to be intentional about the design of expressive robot motions rather than leaving human interpretation of robot motions up to chance. Prior work in HRI has explored using animation techniques in the design of expressive robot motion, e.g., showing forethought and responses to success or failure [64], showing preparation and follow-through in a human-robot jazz performance [22], and shared attention in a music-listening companion robot [18]. Expressiveness has been used to facilitate successful interactions and collaborations with robots. Takayama, et al. (2011), demonstrated how readable robot behaviors can convey rich information about a robots intentions and capability to human observers, as well as change how they subjectively rate the robot [64]. Similarly, Hoffman & Breazeal (2007), showed how adding anticipatory robot motions improved the fluency of human-robot collaborative activities [19,20]. This thesis builds upon these prior works, expanding into the space of curious exploration. Algorithmic approaches to generating expressive robot motions are in development, starting with making robot motions more legible and predictable [8,63] as well as simultaneously generating functional and expressive robot motions [68].

In the current work, we focus upon the expression of curiosity. Some social robots have been developed to express curiosity, e.g., Dragonbots designed to elicit child curiosity [14], Naos designed to play educational games with children [5], and a Wizard-of-Oz Recyclo robot that gave recycling recommendations [35]. These systems have mostly relied upon *verbal* expressions of curiosity, like asking questions or explicitly stating that they are curious about something. They are also designed for learning contexts. In contrast, the current study focuses upon nonverbal, behavioral expressions of curiosity in a professional workplace context. While there are many settings where it could be useful for a robot to express curiosity (e.g., learning contexts [5, 14, 35]), it is unclear if it is always going to be a good idea for robots to express curiosity. For example, when a robot has a specific task that it is supposed to be doing, like delivering mail in an office place, will it be acceptable for the robot to pause its primary task in order to curiously explore other things along the way? Prior work has shown that curious robot behaviors can increase engagement, but can simultaneously hurt trust in the robot [35] and can make people think that the robot is less competent [69]. Additionally, since curiosity based learning robots will always be evolving–theoretically–in their interest, ability, and capability, we need to better understand how to signal that a robot is curious (i.e., capable of being curious) and when it is curious.

Chapter 3

What's in the Box? And Other Things A Curious Working Robot Shouldn't Do

In our first study we examined the impact of a curious robot's behaviors on human perceptions of that robot. We also tested the efficacy of using animation principles to communicate the internal processes of a robot (i.e., AI modeled after curiosity) into behaviors that are innately readable to humans. Finally, we explored whether, and how, curiosity based behaviors would be viewed differently if the robot was deviating from a primary work task or not.

3.1 Study Design

We used a mixed-factorial 4 (curiosity level: none, low, moderate, and high; within-subjects) x 2 (working state: on-duty vs. off-duty; between-subjects) experiment design to explore the following hypotheses:

H1 When a robot expresses more curiosity, people will perceive it as having more curiosity-related traits (e.g., self-motivated to learn, desiring new knowledge).

H2 When a robot expresses curiosity, people will feel more positively toward the robot.

H3 There exists a threshold of acceptability for robot curiosity, beyond which people will see the robot as malfunctioning.

H4 People will be less tolerant of curious robot behaviors when a robot is on-duty as opposed to off-duty.

3.1.1 Stimuli

We developed eight videos of a Mayfield Robotics Kuri operating as a delivery robot in an office setting (See: Figure 3.1). The Kuri was chosen because it is a widely accessible robot designed for interactive social expression that our curious behaviors could be actualized on for in-person follow up experiments. In all of the videos the robot passes by a box in a hallway between cubicles and either ignores it or examines it with various levels of interest (i.e., curiosity). In half of the videos, the scene illustrates that the robot is "on-duty" in the middle of a delivery, and in the other half the robot is presented as "off-duty" with no current tasking.

All scenes were animated with Autodesk Maya, carefully adhering to the joint limits and degrees of freedom of Kuri's hardware. Motion was restricted to realistic base movement, head pan and tilt, and eyelid opening and closing. To fit the narrative we chose, the Kuri model was altered to include a backpack to hold items.

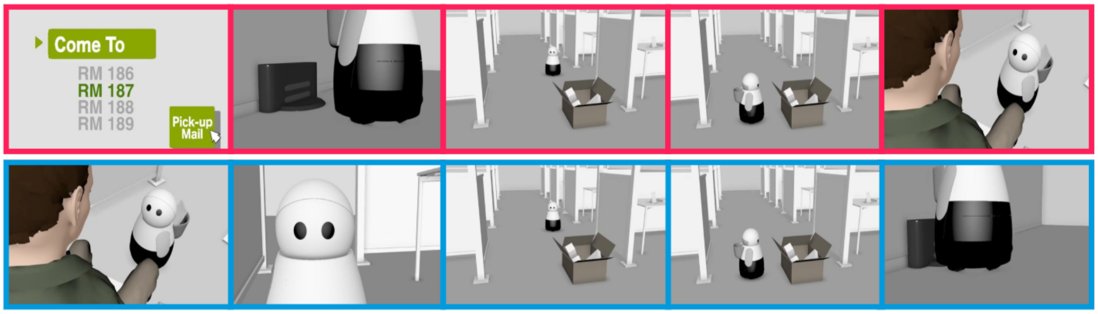


Figure 3.1: Robot's on-duty sequence (top) and off-duty sequence (bottom)
Robot sequence of actions

3.1.1.1 Level of Curious Behavior

We worked with a professional character animator, experienced in motion design for real-world robots, to develop the curiosity behaviors for these studies. Curious behaviors were created, and heightened, by altering the focus, frequency, and duration of the body, head, and eye movements. All of the curiosity behaviors were culled from past research on observable traits of curiosity. These included increasing head movements [44], shifts in gaze and facial expression [51, 52], moving closer to the subject of interest [38–40], and exploring the environment [6, 27, 49].

No Curiosity (L1) = slight head movements and blinks while traveling through the hall, no head rotation toward box, no slowing down, or any acknowledgement box is there, except to move around it.

Low Curiosity (L2) = head pans/tilts looking around while moving through hallway, head rotates toward the box, slows down while passing and head pans/tilts focused on the box, but there is no body rotation towards the box.

Medium Curiosity (L3) = head pans/tilts while moving through hallway, head and body rotate toward the box; robot circles the box with 3 additional pauses, head pans/tilts quickly while looking at the box and surrounding area, before continuing its original route.

High Curiosity (L4) = head pans/tilts while moving through hallway, head and body rotate toward the box; robot circles the box twice with 7 additional pauses, head pans/tilts slowly while looking at the box and surrounding area, before continuing its original route.

Links to all of the stimuli videos can be found in Appendix A.

Because the higher levels of curiosity required more behaviors, the length of the videos were 31, 38, 56, and 126 seconds, respectively, for no, low, medium, and high curiosity. In our animation development and pilot studies, we attempted to hold the time duration constant across conditions, but were unable to do so without introducing confounds to the experiment design (e.g., idle or erratic behaviors). In the end, we made the decision to keep the durations as close together as we could, but to prioritize the expressive curiosity, acknowledging that more curious behavior may inherently take more time to perform. Though a control for time would be ideal past research on curiosity has shown that a major facet of how people determine the intensity of curiosity in others is the amount of time they spend engaged with the subject of interest [38–40, 44]. Our hypothesis that people might be less patient with robot curiosity in on-duty situations (H4) included consideration that more curious robots might take more time to explore rather than doing their task at hand.

3.1.1.2 Working state: On-Duty Vs. Off-Duty

In the *on-duty* videos, a human is shown looking at a letter and then using an application on their computer to ask the robot to come retrieve the mail in their office. The robot is shown activating and leaving its charging station. Next the robot enters a cubicle hallway with a box and performs one of the levels of curious behaviors (see above) focused on the box before leaving. Finally, the robot arrives at the humans desk and the letters are loaded into its backpack.

The *off-duty* videos begin with a robot arriving at a person's cubicle carrying letters in its backpack. The robot is seen completing its delivery task and a screen is shown indicating the robot is off-duty and intending to return home to its charger. Next the robot enters a cubicle hallway with a box and performs one of the levels of curiosity behaviors focused on the box before leaving. In the final shot the robot arrives at its charging station and parks itself.

The videos were designed to mirror each other closely in order to minimize possible context-based confounds. Most of the shots used in each set of videos are nearly identical with the order they are shown simply reversed and minor additions made to establish the narrative on-duty vs. off-duty distinction. The hallway portions, where the robot behaves curiously, are identical in both sets of videos. See Figure 3.1. For more details, see the visual stimuli linked in Appendix A.

3.1.2 Method

Participants were randomly assigned to one of the between-subject conditions–on-duty or off-duty. Each participant was shown all four levels of robot curiosity in a randomized order. The within-subjects conditions were presented in a randomized order and potential order effects were checked for in our data analyses. Past research on curiosity has generally used bipolar treatment variables (i.e., curious vs. not curious) as such we were concerned that response to shifts in curious behavior may be fairly subtle. As such we chose to operationalize curiosity as a within subjects variable since we wanted to reduce noise in the data as much as possible. Similarly, we chose to treat working state as a between subjects variable because we were concerned with order effects as well as participant fatigue if people had to watch eight stimuli videos, rather than four.

3.1.2.1 Participants

Respondents (N=30) were recruited from Amazon Mechanical Turk. Each participant received two U.S. dollars as a token of thanks for their time. It took approximately 10 minutes to complete the study. From an original pool of 31, one participant was removed for having a response time significantly lower than the total duration of the videos they were asked to watch. There were 17 males and 13 females (self-reported) with ages ranging from 25 to 69 years ($M = 38.9$, $SD = 10.2$). Education levels of respondents included completion of high-school ($n = 5$), some college ($n = 4$), associates degrees ($n = 4$), bachelors degrees ($n = 14$), and masters degrees ($n = 3$). 22 out of 30 participants reported pet ownership experience. In terms of reported experience with robots; 22 had none, 3 a little, 2 a moderate amount, 1 a great deal, and 2 preferred not to answer.

3.1.2.2 Procedure

Participants were informed that they would watch several short video clips, featuring robots, and give their opinions on the robots they saw. Those participants who accepted the task were directed to a survey where they were asked to read and agree to a digital consent form. In the consent form participants were also informed that the task included several attention checks and that if they failed them the work would be rejected. Participants were then randomly assigned to the on-duty or off-duty conditions. Each participant saw videos that depicted four different curiosity levels (none, low, medium, high), randomized for order. They were asked to watch each video in full screen mode. After each video, the participant responded to a questionnaire about the robot they just saw. Each questionnaire consisted of fourteen items presented on a 7-point Likert scale of agreement followed by one open-ended item that asked them to describe the robot they saw. Three items were reverse coded and question order was always randomized.

We designed the survey items to gauge readability of our curiosity animations as well as assess participants' general perceptions of the robot. We chose questions focused on perceived curiosity based on common tangible/observable factors of curiosity (i.e., pertaining to curiosity as behavior, not mental state). Similarly, we picked items regarding general perception of robots in order to capture common aspects of HRI research (e.g., trust, likability, intelligence). We also included an attention check for our working state between-subjects variable. After the final video, we asked participants to select from four options to describe what the robot was doing when it passed the box in the hall. Finally, we asked participants to fill out a brief demographics questionnaire about their age, gender identity, education level, pet ownership, and experience with robots.

3.2 Quantitative Data Analysis

Below we review the details of our data analysis approach and present results for the quantitative measures we collected (i.e., likert-type questionnaire items). A similar review of how our qualitative data was examined is discussed further down in: Section 3.3).

3.2.1 Order Effects and Manipulation Checks

Since curiosity level was a within-subjects independent variable in this study, there was a risk that we would see order effects in the data (i.e., having seen the first condition would affect responses to the subsequent conditions). To address this concern, we re-structured the data in terms of first, second, third, and fourth curiosity condition presented to each participant. We then ran a mixed ANOVA, using order of curiosity stimuli presented (first, second, third, fourth) as a within-subjects independent variable and working state as a between-subjects independent variable. In running analyses for all of the dependent variables, we did not find any statistically significant effects of order of stimuli presentation in these data.

To see whether participants viewed our different levels of curious expression as intended, we ran a repeated measures ANOVA, using level of behavioral curiosity as the within-subjects independent variable and participant ratings of the survey item "the robot is curious" as the dependent variable. There was a significant main effect of the robot's behavioral curiosity level on participants' perceived curiosity of the robot $F(3, 84)=17.06$, $p=.001 < .05$, $\eta_p^2=.38$. There was no effect of working state (on vs. off-duty) on perceived curiosity. We ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant differences between levels 1 vs. 2 $F(1,28)=9.34$, $p=.005$, $\eta_p^2=.25$ and levels 2 vs. 3 $F(1,28)=18.37$, $p=.001$, $\eta_p^2=.4$. See Figure 3.2.

3.2.2 Hypothesis Testing

We ran a mixed ANOVA, using level of behavioral curiosity as a within-subjects independent variable and working state as a between-subjects independent variable. The mixed ANOVAs were used to predict each of our three constructed components. We also ran mixed ANOVAs to predict our remaining independent survey items (responsibility, focus, and competence), and the manipulation check of perceived curiosity. Where there were significant effects for curiosity level, we also ran repeated contrasts between the four levels of curiosity.

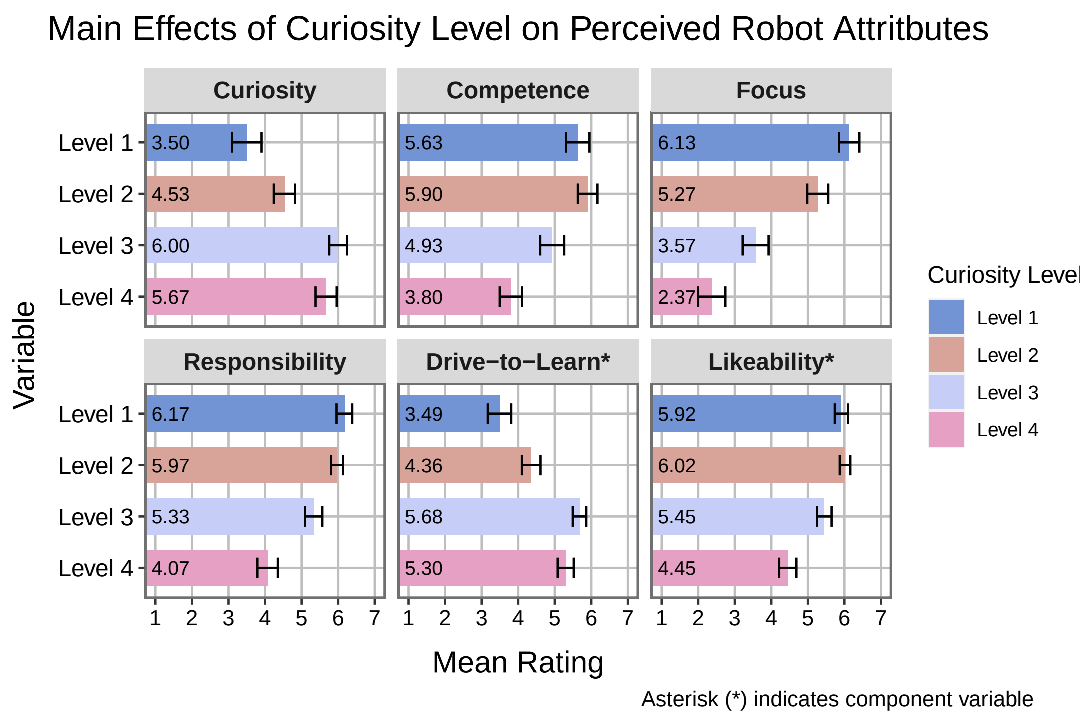


Figure 3.2: (Top Left) Manipulation check of robot curiosity behaviors demonstrates that curiosity levels impact the perceived curiosity. (Top Center and Right) Main effects of robot curiosity on ratings of competence, focus, and (Bottom left) responsibility. (Bottom Center and Right) Main effects of curiosity on perceived drive-to-learn and likability components.

3.2.3 Results

Different sets of our survey items we saw as relating the broader concepts we set out to explore in this work. Namely, how curiosity might change perceptions of the robot as having cognition or thought, being a likable social character, seeming self-sufficient or not needing human oversight, and as a capable or competent worker. We chose to extract and examine these higher level combinations of survey responses for the sake of easier interpretation of our largely exploratory data analysis [28]. First, we performed principal component analysis (PCA) at each of the four levels of our repeated measure (level of curious behavior) to create our primary dependent variables. We based our PCA upon four conceptual groupings of perceived *Drive to Learn* (the robot learns on its own, would explore, and desires knowledge) *Likability* (the robot is likable, acceptable, trustworthy, and approachable), *Self-Sufficiency* (the robot is independent, intelligent, and would take initiative on new tasks), and *Competence* (the robot is focused, competent, and responsible) of the robot. To determine which items were included in each component, we used the recommended component loading minimum of 0.4 [48, 62]. Of our conceptual components only *Competence* failed to hold together at all levels of our repeated measure, so we analyzed the responsible, distracted (reverse coded), and competent (reverse coded) items individually. For mean and standard error values of all components and items see: Table 1. Appendix A. Finally, we then conducted reliability tests for each of our new components, which all returned acceptable Cronbach's alpha ($\alpha > 0.7$) at each level of our repeated measure (For details see: Appendix B.2).

3.2.3.1 Perceived Drive to Learn

The perceived drive to learn component was calculated as the average score across the items: This robot would learn on its own, explore its surroundings, and has a desire for new knowledge.

There was a significant main effect of the robot's behavioral curiosity level on participants' perception of the robot's drive to learn $F(3,84)$=29.04, p=.001, η_p^2=.52. See Figure 3.2. These findings support H1. We did not observe a main effect of working state on perceived curiosity of the robot or an interaction effect between the two

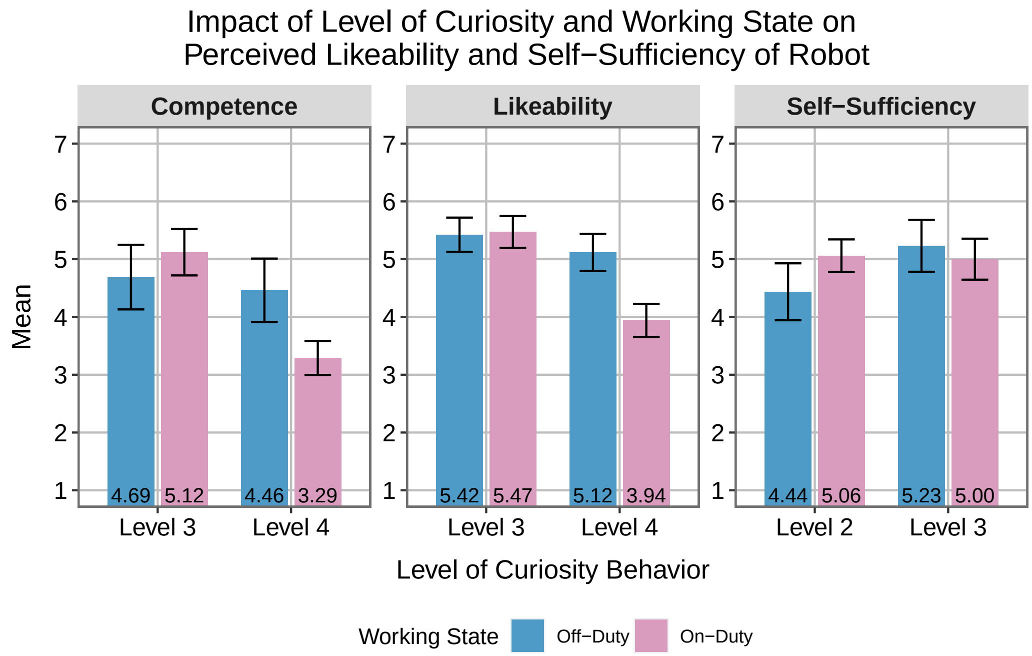


Figure 3.3: Effects of robot's curiosity and working state (on-duty vs. off-duty) on competence, likability, (levels 3 vs. 4), and self-sufficiency (levels 2 vs. 3).

variables.

Following up on the significant effects, we ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant differences between levels 1 vs. 2 $F(1{,}28)$=14.48, p=.001, η_p^2 = .36 and levels 2 vs. 3 $F(1{,}28)$ = 33.98, p=.001, η_p^2 = .55.

3.2.3.2 Robot Likability

The robot likability factor was calculated as the average score across four items: This robot is likable, acceptable, trustworthy, and approachable.

We found a significant main effect of the robot's curiosity behaviors on ratings of robot likability F(3,84)=35.6, p=.001, partial η^2=.56. See Figure 3.2. People liked the robot less as it behaved with more curiosity. This finding does not support H2. Rather, it shows support for the *opposite* of H2, as curiosity increased likability went down. In other words, as a robot's curiosity increased, its likability went down. We did not observe a statistically significant effect of working state on likability. We found a significant interaction effect between curiosity level and working state, F(3,84)=6.20, p=.001, η_p^2=.18. See Figure 3.3.

Following up on the significant effects, we ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference in likability of the robot between curiosity levels 2 vs. 3 F(1,28)=19.03, p=.001, η_p^2=.41, and levels 3 vs. 4 F(1,28)=32.19, p=.001, η_p^2=.54. We did not find a difference between levels 1 vs. 2. We also found a significant interaction between levels 3 vs. 4, F(1,28)=14.23, p=.001, η_p^2=.34. See Figure 3.3. People liked the robot least when it was on-duty, being extremely curious; this finding supports H4.

3.2.3.3 Perceived Self-Sufficiency

The self-sufficiency component was calculated as the average score across the items: This robot is independent, intelligent, and would take initiative to perform tasks.

There was a significant interaction effect between curiosity level and working state upon perceived robot self-sufficiency, F(3,84)=3.87, $p<.05$, $\eta_p^2 = .12$. We did not see statistically significant main effects. Following up on the interaction effects, we ran repeated contrasts between each of the four levels of curiosity. In that analysis, we observed a significant interaction between levels 2 vs. 3, $F(1{,}28)$=6.53, p<.05, $\eta_p^2 = .15$. See Figure 3.3. We did not find significant differences between Levels 1 vs. 2 or 3 vs. 4.

3.2.3.4 Focus, Competence, and Responsibility (single items)

Because the Competence component was not found to be reliable, we analyzed each of these three as single-question items.

We found a significant main effect of curiosity behaviors on the **perceived focus** of the robot $F(3{,}84) = 34.11$, p=.001, $\eta_p^2 = 72$. There was no main effect of working state on perceived focus or interaction effects present. We ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between all levels of curiosity behavior: Level 1 vs. 2 $F(1{,}28) = 4.62$, p=.04, $\eta_p^2 = .14$, level 2 vs. 3 $F(1{,}28) = 19.68$, p=.001, $\eta_p^2 = .41$, and levels 3 vs. 4 $F(1{,}28) = 12.33$, p=.002, $\eta_p^2 = .31$. See Figure 3.2.

We found a significant main effect of curiosity behaviors on the **perceived competence** of the robot $F(3{,}84)$=15.85, p= .001, $\eta_p^2 = .36$. We also found a significant interaction effect $F(3, 84)$=3.76, p=.014, $\eta_p^2 = .12$. See Figure 3.3. There was no main effect of working state on perceived focus. With repeated contrasts on each of the four levels of curiosity behaviors we found significant difference at levels 2 vs. 3 $F(1{,}28)$=8.84, p=.006, η_p^2=.24, and levels 3 vs. 4 $F(1{,}28)$=14.32, p=.001, η_p^2=.34. See Figure 3.2. This finding supports the *opposite* of what we predicted in H2. It supports H3. For the interaction effect, we found significant difference between level 3 vs. 4 $F(1{,}28)$=8.61, p=.007, η_p^2=.24. People thought the most curious robot was the least competent, especially when it was on-duty; this finding supports H4.

We found a significant main effect of curiosity behaviors on the **perceived responsibility** of the robot $F(3,84)$=22.94, p=.001, η_p^2=.45. There was no main effect of working state on responsibility or interaction effects. Repeated contrasts on each of the four levels of curiosity behaviors found significant differences between levels 2 vs. 3 $F(1,28)$=8.44, p=.007, η_p^2=.23, and levels 3 vs. 4 $F(1,28)$=24.0, p=.001, η_p^2=.46. See Figure 3.2. These findings do not support H2, but they do support H3.

3.3 Qualitative Data Analysis

In our analysis of survey items, we saw a generally negative shift in participants' ratings of the robot–aside from increased ratings of its cognitive ability–as curiosity behaviors went up. After each video (i.e., each level of robot curiosity) participants responded to the open-ended question: "How would you describe the robot you just saw?" Since each participant provided their own perspective on the robot, in the form of an open-ended question after each level, we decided to examine their responses more closely in hopes of identifying, not just the underlying symptom of the negative ratings, but also the potential cause. We examined these responses through both sentiment analysis of emotional tone and inductive thematic coding. Our approach and findings for each of these methods are presented below.

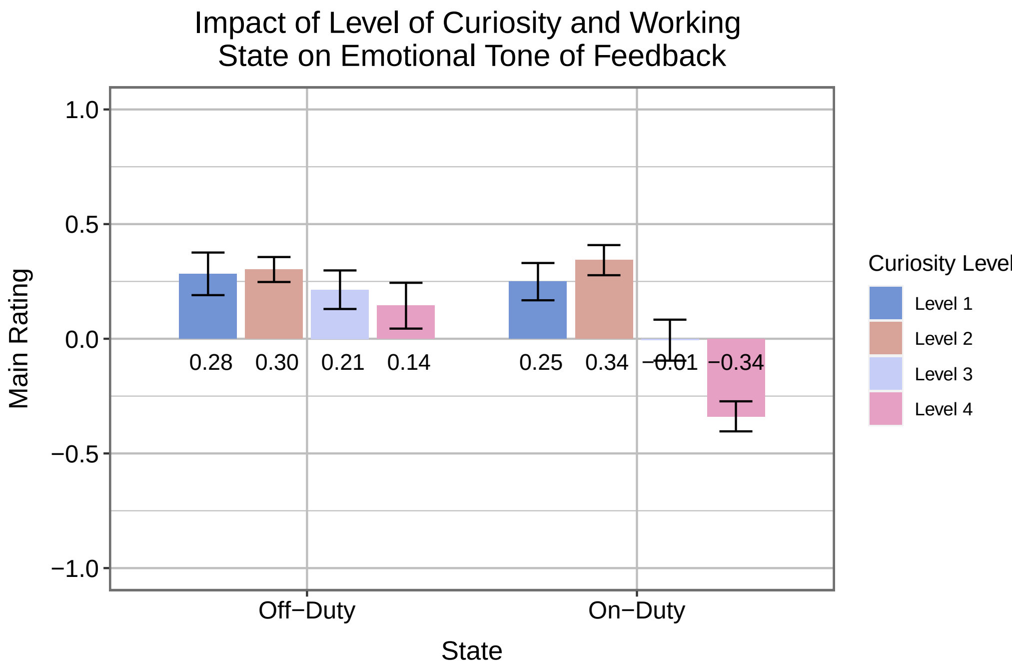


Figure 3.4: Robot curiosity behavior and working state (off-duty vs. on-duty) impacted emotional tone of participant open-ended responses. Significant change occurred at the highest level of curiosity (level 3 vs. 4).

3.3.1 Sentiment Analysis

First we used sentiment analysis in order to look for emotional changes in participants' written feedback across conditions. We conducted sentiment analysis on participants' open ended responses using VADER (see: VADER for open source access). VADER generates a sentiment score that can be interpreted as positive, negative, or neutral using a 1 to -1 scale, which provides a reliable metric for emotional valence in short text entries [25]. Next we ran a mixed ANOVA, using level of curiosity as a within-subjects independent variable and a robot's working state (On-Duty vs. Off-Duty) as a between-subjects independent variable to predict compound emotional valence of participant responses.

We found significant main effects of curiosity, $F(3{,}84)$=52.76, p<.001, η_p^2=.28, and a robot's working state, $F(1{,}28)$=8.68, p<.01, η_p^2=.24, on the compound emotional tone of participant descriptions of the robot. We also captured a significant interaction effect between our two main variables, $F(3{,}84)$=4.26, p=.01, η_p^2=.13, on the emotional tone of their responses. Next, we ran repeated contrasts on each of our independent variables. Here we found significant difference between levels 2 vs. 3, $F(1{,}28)$=10.54, p<.005, η_p^2=.27, and levels 3 vs. 4, $F(1{,}28)$=5.77, p<.05, η_p^2=.17. We did not find a significant difference between levels 1 vs. 2. Our repeated contrasts did not reveal where specifically a significant difference interaction occurred between any levels of curious behavior and the working state of the robot.

3.3.2 Thematic Coding

We took an inductive approach to develop a thematic code-book and tag participant responses describing the robots they saw [65]. In effect, we spent substantial time reading and discussing our participant comments in order to attune with the data, allowing patterns and meaning to emerge [7].

We began by reviewing a randomly selected subset of 20% of participant ($n = 6$) responses ($n = 24$), using a line by line approach wherein code(s) were created and applied to each response separately. Once we reached saturation (i.e., there were no new emergent codes), we reviewed the initial codes and organized them into tentative themes. Descriptions of meaning and examples were assigned to each code. To assess the reliability of the code book, we performed inter-reliability analyses on the data, using Cohen's κ. The primary coder coded the full data set and the reliability coder blind-coded responses from a new randomly selected subset of six participants. We calculated Cohen's κ for each set of codes to assess the consensus between the two coders. When the reliability was not sufficiently high ($\kappa < .70$), disagreements about the meaning of codes were discussed and resolved by refining the code book and re-coding the data, using the updated code book. This inductive iteration process was conducted a total of three times, using a new random subset of responses from six participants. Having demonstrated reliability, the primary coder performed a final coding pass on the full data set. Our final code book (See: Appendix C.1) consisted of four pairs of mutually exclusive codes. Each set had high inter-rater reliability: Focused/unfocused ($\kappa = .804$), reliable/unreliable (κ = .931), obedient/disobedient (κ = .714), and positive/negative cognitive traits (κ = .852). Responses were coded with one of the mutually exclusive items from each set or with "none."

3.3.2.1 Focus, Reliability, Obedience, and Cognition

Participant responses often included a critical reflection on **how** the robot went about its business, either delivering mail or returning to the charger. We called this set "focus," which distinguished between two codes, **focused vs. unfocused**. This arose when participants described a deviation of the robot's attention in a way that they viewed as appropriate or not. For example:

> Unacceptable. Clearly was distracted if not preoccupied with the box and I was surprised it went to get the letter at all. –P6 (L4/On-Duty)

Participants also reflected on how confident they were in the robot's performance more broadly. We coded this set as **reliable vs. unreliable**. This arose when participants reflected on whether or not the robot could be depended upon to complete tasks, not only in this particular setting, but overall. The pair of responses below highlights how the participant's confidence in the robot differed between levels of curiosity.

> It seems curious to its surroundings but still focused on the task of returning home. It would be able to overcome obstacles it encountered. –P15 (L3/Off-Duty)

> Faulty and undependable. Untrustworthy to do what was needed. I think it would get into trouble because it does not obey commands. –P15 (L4/Off-Duty)

These responses also raised the theme of obedience, distinguished by **obedience vs. disobedience**. This set resulted from participants attributing a reason **why** the robot acted certain ways in relation to the task at hand. They might describe the robot as obedient by listening to instructions, obeying commands, or following its programming properly. Conversely, they would describe the robot as disobedient by choosing not to follow instructions or malfunctioning in some way.

Our final set of thematic codes came from participants attributing different mental capacities to the robot as **positive vs. negative cognitive traits**. Positive was coded when participants discussed the robot possessing useful or higher cognitive traits, like the robot being intelligent, being interested in learning, or wanting to investigate its surroundings. Negative was used when participants applied negative or lesser cognitive traits to the robot. For example:

> Extremely slow and easily distracted. I would call it "ditzy". It wasted way too much time meandering around the box. –P8 (L4/On-Duty)

While the quantitative results of Study 1 showed us how people rated the robot in response to closed-ended questions, these qualitative results enabled us to dig deeper into the potential reasons for why they perceived the robot more positive or negatively when displaying different levels of curious behavior.

We ran a χ^2 analysis to assess whether the frequency of each code differed across each of the levels of robot curiosity. This enabled us to ask questions like: Did people mention obedience more or less often when the robot behaviorally expressed greater amounts of curiosity? The descriptive results are presented in Figure 3.5.

Across the four levels of curiosity, we found statistically significant differences in the number of people who mentioned the robot being; focused (χ^2=17.3, p=.0006), unfocused (χ^2=27.0, p=.0000), reliable (χ^2=14.2, p=.0027), unreliable (χ^2=19.9, p = .0002), obedient (χ^2=15.7, p=.0013), disobedient (χ^2=14.1, p=.0027), having positive cognitive traits (χ^2=8.18, p=.0424), and having negative cognitive traits (χ^2=14.6, p=.0022). Next we ran pair-wise χ^2 tests to identify between which levels of curious behavior significant change occurred. We found significant differences in the number of people who mentioned the robot being; unfocused levels 2 vs. 3 (χ^2=4.5, p=.0339), reliable levels 3 vs 4. (χ^2=7.0, p=.0082), unreliable levels 3 vs 4. (χ^2=4.45, p=.0348), obedient levels 2 vs. 3 (χ^2=5.44, p=.0196), and having positive cognitive traits levels 1 vs. 2 (χ^2=4.76, p=.0290). We also ran χ^2 analysis to assess whether the frequency of each code differed between working state conditions (on-duty vs. off-duty). Between working state conditions we found significant differences in the number of people who mentioned the robot being; unfocused (χ^2=6.0, p=.0143), unreliable (χ^2=4.45, p = .0348), and having negative cognitive traits (χ^2=9.00, p=.0027).

The trends we observed in these qualitative data include: (1) as curiosity level went up, people talked more about the robot being unfocused, unreliable, disobedient, and having negative cognitive traits and (2) as curiosity level went down, people talked more about the robot being focused, reliable, obedient, and having positive cognitive traits. Additionally, we saw that when the robot was framed as busy working (on-duty) people talked more about it being unfocused, unreliable, and having negative cognitive traits.

These open-ended responses gave us a better sense for how people made sense of the robot's behaviors, using their own words. One caveat here is that they gave

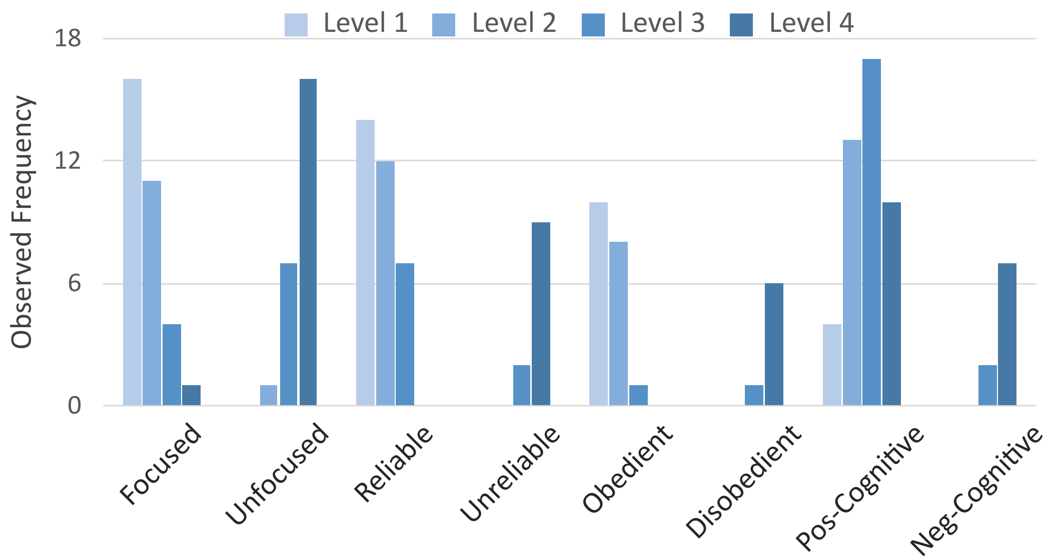


Figure 3.5: Frequency distribution of thematic codes at each level of robot curiosity.

these open-ended responses after they answered the closed-ended questionnaire, which included items that asked about distraction and reliability.

The two sets of codes that do not overlap with items in the closed-ended questions were obedience vs. disobedience and positive vs. negative cognitive traits. People commented upon the less curious robots as being obedient and the more curious robots as being disobedient, even though we did not ask them about obedience anywhere else in the study protocol. This suggests that people are attributing intentionality to the robot's decision to obey or disobey the requests of the person in the scenario. Also, we observed a trend of people making more comments about the robot's cognitive traits as its behaviors became increasingly curious. However, at the highest level of curious behaviors, people tended to comment upon more negative cognitive traits (e.g., being ditzy).

3.4 Summary of Study 1 Findings

Study 1 demonstrated that realistic robot behaviors can be used to express robot curiosity in a way that untrained end-users can readily interpret. We sought to test hypotheses about what other traits people would associate with more curious robot behaviors (H1), that people would feel more positively about more curious robots (H2), whether there might be such a thing as having too much robot curiosity (H3), and whether people would be even less accepting of robot curiosity when the robot is supposed to be doing an assigned work task (H4). These data supported Hypothesis 1, 3, and 4, but we did not find support for Hypothesis 2. In fact, we found evidence for the opposite of Hypothesis 2, that people liked the robot less the more curious it was.

We also found that people indeed perceived the more curious robot behaviors as the robot being more curious (manipulation check) and having a stronger drive-to-learn, supporting H1. People felt more negatively toward the more curious robot behaviors, including liking it less and rating it as being less focused, less competent, and less responsible, opposing H2 and supporting H3. The significant drops in these ratings largely fell between the higher levels of curiosity (Levels 2 to 3 and 3 to 4), not between Levels 1 and 2, which suggests that a low level of curious behavior might be acceptable. When looking at on-duty vs. off-duty robot settings, we found that people felt most negatively about the on-duty robot behaving with extremely high curiosity, supporting H4. The semantic analysis of participants' open-ended responses, using VADER [25], provided complimentary evidence to these findings. Here, as the robot behaved with more curiosity the emotional tone of responses became significantly more negative and, again, was exacerbated at the higher levels of curiosity when the robot was framed as on-duty.

Having seen how a robot's curiosity and working state impacted participant ratings we turned to their qualitative responses in an effort to also understand why this happened. In our examination of open-ended responses, we noticed a trend of people commenting on the robot's cognitive traits more as it behaved in more curious ways, but those comments became more negative at the more extreme levels of curious robot behavior. This trend, in the open-ended responses, looked much more like what we had expected to see, in the survey items, when we formulated Hypotheses 2 and 3, increasing positivity toward more curiosity, but falling off at a threshold between Levels 3 and 4. Additionally, people spontaneously raised issues of robot obedience vs. disobedience, which indicates they perceived the robot as having intentionality. Moreover, this seemed to be tied in to a perceived violation of a social contract between the robot and the human counterpart wherein the more curious robots, seeming to be endowed with greater mental capacity (but not necessarily ability), were scrutinized more severely for deviations.

Chapter 4

What to Expect When You're Expecting a (Curious, Learning, or Autonomous) Robot

In the prior study we found that participants responded differently to a robot as it's curious behavior increased (e.g., higher ratings of mental ability but lower ratings of likability and competence). Additionally, we saw that at high levels of curious behavior people became sensitive to the perceived working state of the robot, rating it more negatively when it was framed as being on-duty. In this follow-up study, we examined whether priming participants to have different expectations of the robot (i.e., being curious, learning, or autonomous) would alter their perceptions of increasingly curious behavior. We also examined how well a robot's behavior matched participants' expectations based on how they were primed. Finally, following each treatment level, participants provided open ended descriptions of the robots' behavior and explanations about what they liked/disliked, and why.

4.1 Study 2 Design

We used a mixed-factorial 4 (curiosity level: none, low, moderate, and high; within-subjects) x 3 (expectation priming: curious vs. learning vs. autonomous; between-subjects) experiment design to test the following hypotheses:

H1 People hold distinct mental models for robots (how it should behave) depending on the cognitive capacities they believe it has.

H2 People primed to expect a **curious robot** will rate the robot as more effectively matching their expectations as the level of curious behavior goes up.

H3 People primed to expect a **learning robot** will rate the robot as more effectively matching their expectations at the lower levels of curious behavior but lower at the highest level of curiosity.

H4 People primed to expect an **autonomous robot** will rate the robot as more effectively matching their expectations when it exhibits little or no curious behavior and not when it exhibits greater amounts of curious behavior.

In particular we were interested in the hypothesized difference between a curious and a learning robot for two reasons. First, in our review of literature on computational models of the human mind, broadly, and curiosity, specifically, we found a tendency to use, interchangeably, any synonym of the same. Yet, we saw compelling arguments–often in the same works–that concepts of human thinking, learning, curiosity, etc., are quite distinct in both human function and how they are operationalized for computational purposes. This brings us to the second point: The meaning of learning is somewhat more bounded by potential outcomes (i.e., something becomes learned) whereas curiosity is, comparatively, self-sustaining because it is a motivational state (i.e., something becomes curious) without a specific end point.

We were also interested in whether priming participants to expect a robot with specific mental attributes would offset any of the negative trends we observed in the prior study (i.e., as the robot became more curious it was seen as less competent, focused, responsible, and likable). As such we posed the following research question:

- Will priming people to anticipate specific types of internal mental states in a robot (i.e., curious, learning, or autonomous) offset negative reactions to increasing curiosity behaviors observed in the previous study?

4.1.1 Method

Participants were randomly assigned to one of the three between-subject conditions: curious robot, learning robot, or autonomous robot. Each participant was shown all four levels of robot curiosity in a randomized order. We used the same four "on-duty" videos as the previous study (See: Appendix A) which depict a Mayfield Robotics Kuri operating as a delivery robot in an office setting. The only changes we made to the video stimuli in the present study was the inclusion of a five second title card added to each video stating: "This is a (learning, curious, or autonomous) robot that delivers mail in an office." For additional details on the curious behavior design process and its relation to past work on human curiosity see: Section 3.1.1.

4.1.1.1 Participants

Respondents (N=64) were recruited from Amazon Mechanical Turk. Each participant received two U.S. dollars as a token of thanks for their time. It took approximately 10 minutes to complete the study. From an original pool of 69, five participants were removed for having a response time significantly lower than the total duration of the videos they were asked to watch and/or failing attention check via straight-lining behavior. There were 39 males, 23 females, 1 non-binary, and 1 person who preferred not to answer (self-reported) with ages ranging from 26 to 61 years (M = 38.98, SD = 8.9). Education levels of respondents included completion of high-school (n = 4), some college (n = 14), associates degrees (n = 7), bachelors degrees (n = 31), masters degrees (n = 7), and professional degrees (n = 1). In terms of reported experience with robots; 55 people had none, 3 a little, 3 a moderate amount, 2 lots, and 1 preferred not to answer.

4.1.1.2 Procedure

In this experiment we used the same general protocol for recruitment and distribution as the previous study. For more details see: 3.1.2.2. We randomly assigned participants, and counter-balanced, to one of three between-subjects conditions (curious robot n = 21, learning robot n = 21, and autonomous robot n = 22). Depending on condition, we told participants they would see videos of a “curious”, “learning”, or “autonomous” robot working in an office. Then we requested they write record their expectations, with an open ended item, of how a curious robot, a learning robot, or an autonomous robot should behave (depending on condition).

Next, participants saw videos depicting four different curiosity levels (none, low, medium, high), randomized for order. Each video began with a five second title card reiterating that they will see a curious, learning, or autonomous robot working in an office delivering mail. They were asked to watch each video in full screen mode. After each video, participants responded to a questionnaire about the robot they just saw. Each questionnaire consisted of nineteen items presented on a 7-point Likert scale of agreement. We used the same set of 14 survey items from our first study (See: 3.1.2.2) and added an additional five, explained below. In the prior study, qualitative analysis of participants’ open ended responses, indicated that people were often concerned with whether the robot was reliable, adaptable, or obedient, and whether it would follow instructions. Four new survey items were added to capture these factors. The original survey included questions pertaining to the robot as a curious or learning agent. We added one final item to gauge perceptions of the robot as an autonomous agent. Four items were reverse coded and question order was always randomized.

Following the above items we also asked participants to rate, on a single 5-point Likert type scale, how well the robot they saw matched the expectations they recorded at the beginning of the study. Next, we asked participants to fill out two open ended items describing the robots’ behavior and what they liked/disliked about the robot.

The 7-point scale of agreement items, the 5-point expectation matching item, and the two open ended responses were repeated following each of the four videos.

Finally, we asked participants to complete a brief demographics questionnaire for their age, gender identity, education level, and experience with robots.

4.2 Quantitative Data Analysis

As in our first study, we began our analysis by pre-processing our data. We performed principal component analysis (PCA), scale reliability testing, and component construction at each of the four levels of curious robot behavior. We identified three components which loaded together, at all four levels of our repeated measure, with acceptable loading scores greater than .40 [48,62]. Then we constructed three new components with the mean scores of related items from our Likert type scales of agreement. First, was the robot as a "thinking agent" which was made of items pertaining to the robot; desiring knowledge, learning on its own, being independent, intelligent, and curious. Second, was the robot as a "working agent" which was constructed from items of the robot being; focused, competent, obedient, unobtrusive, acceptable, and following instructions. Third, was the robot as a "social agent" which was built from items relating to the robot being; trustworthy, likable, responsible, and approachable. Next, we conducted reliability testing on each of the three component sets, which all returned acceptable Cronbach's alphas ($\alpha > 0.7$) at each level of our repeated measure (For details see: Appendix D.1).

There were four individual items that were strongly independent of all others during PCA. These remaining items related to the robot; being autonomous, adaptable, exploring on its own, and taking initiative to perform new tasks.

4.2.1 Results

We ran a mixed ANOVA, using level of curiosity as a within-subjects independent variable and priming expectations of a robot's cognitive state (Curious vs. Learning vs. Autonomous) as a between-subjects independent variable to predict how well the robots' behaviors matched participants' expectations. We found a significant main effect of the robots' curiosity behaviors, $F(3,183)=7.03$, $p<.001$, $\eta_p^2=.10$, on how well the robots' behaviors matched expectations of participants. To identify where changes occurred we ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between levels 1 vs. 2, $F(1,61)=23.22$, $p<.001$, $\eta_p^2=.28$, and levels 3 vs. 4, $F(1,61)=17.20$, $p<.001$, $\eta_p^2=.22$. Our repeated contrasts did not reveal a difference between levels 2 vs. 3. There was also a main effect of the anticipated cognitive state (Curious vs. Learning vs. Autonomous), $F(2,61)=2.15$, $p<.05$, $\eta_p^2=.12$, on expectation matching. Additionally, we captured a significant interaction effect between the robots' behavioral curiosity and anticipated cognitive state, $F(6,183)=15.93$, $p<.001$, $\eta_p^2=.10$, and the robot matching participant expectations. Our contrasts revealed a significant difference in interaction at all levels of curiosity: Levels 1 vs. 2, $F(2,61)=8.48$, $p=.001$, $\eta_p^2=.22$; levels 2 vs. 3, $F(2,61)=19.90$, $p<.001$, $\eta_p^2=.31$; and levels 3 vs. 4, $F(2,61)=3.56$, $p<.05$, $\eta_p^2=.11$. See Figure 4.1.

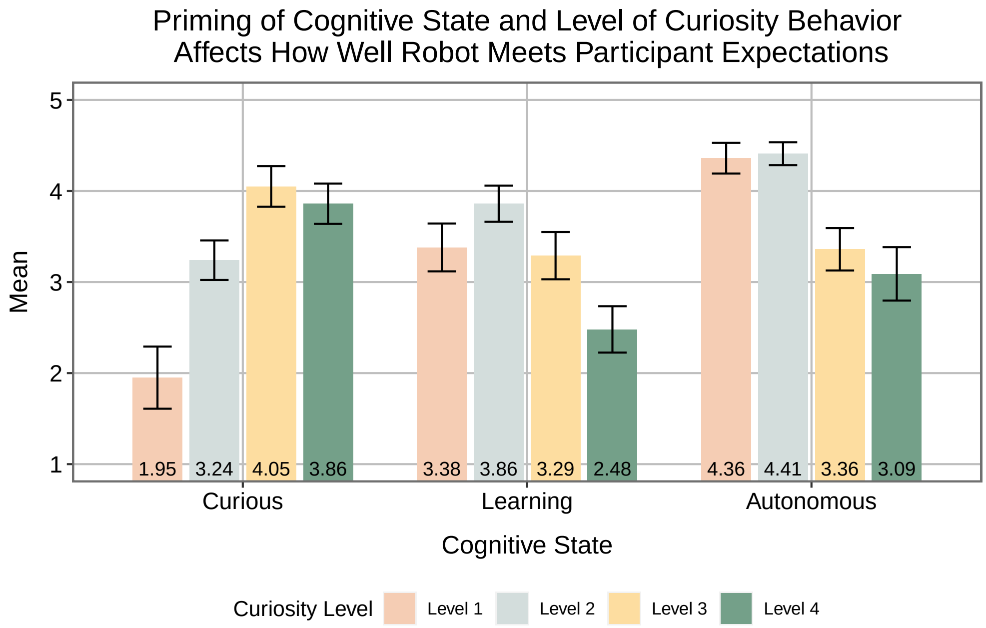


Figure 4.1: Impact of robot's curiosity and anticipated cognitive state (Curious vs. Learning vs. Autonomous) on how well the robot met expectations of participants.

We ran mixed ANOVAs using level of curiosity as a within-subjects independent variable and priming expectations of a robot's cognitive state (Curious vs. Learning vs. Autonomous) as a between-subjects independent variable to predict ratings of the robot on each of our three constructed components (i.e.,thinking agent, working agent, and social agent) separately. We found a significant main effect of a robot's curiosity behavior, $F(3{,}183)=50.09$, $p<.001$, $\eta_p^2=.45$, on the "thinking agent" component. To identify where the change in ratings of the robot as a thinking agent occurred, we ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between levels 1 vs. 2, $F(1{,}61)=62.47$, $p<.001$, $\eta_p^2=.51$, and levels 2 vs. 3, $F(1{,}61)=20.60$, $p<.001$, $\eta_p^2=.25$. We did not observe a significant difference between levels 3 vs. 4. Additionally, we found a significant main effect of the robots' curiosity behaviors, $F(3,183)=96.49$, $p<.001$, $\eta_p^2=.61$, on ratings of the "working agent" component. To identify where significant change in ratings of the robot as a working agent occurred, we ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between levels 1 vs. 2, $F(1,61)=30.33$, $p<.001$, $\eta_p^2=.33$, levels 2 vs. 3, $F(1,61)=55.84$, $p<.001$, $\eta_p^2=.48$, and levels 3 vs. 4, $F(1,61)=28.86$, $p<.001$, $\eta_p^2=.32$. We also found a significant main effect of the robots' curiosity behaviors, $F(3,183)=27.68$, $p<.001$, $\eta_p^2=.31$, on ratings of the "social agent" component. To identify where the change in ratings of the robot as a social agent occurred, we again ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between levels 2 vs. 3, $F(1,61)=12.63$, $p=.001$, $\eta_p^2=.17$, and levels 3 vs. 4, $F(1,61)=26.98$, $p<.001$, $\eta_p^2=.31$, but not between levels 1 vs. 2. We did not observe a main effect for any of the robot cognitive state primes or interaction effects for participant ratings of the robot as a thinking, working, or social agent. See: Figure 4.2.

Finally, we ran mixed ANOVAs using level of curiosity as a within-subjects independent variable and priming expectations of a robot's cognitive state (Curious vs. Learning vs. Autonomous) as a between-subjects independent variable to predict ratings of the robot on our four remaining individual items (i.e.,. adaptable, likely to explore, taking initiative on new tasks, and autonomous). We found a significant main effect of the robots' curiosity behaviors, $F(3,183)=7.33$, $p<.001$, $\eta_p^2=.11$, on ratings of the robots' adaptability. To identify where the change in perceived robot exploration occurred, we ran repeated contrasts on each of the four levels of curiosity behaviors. Here

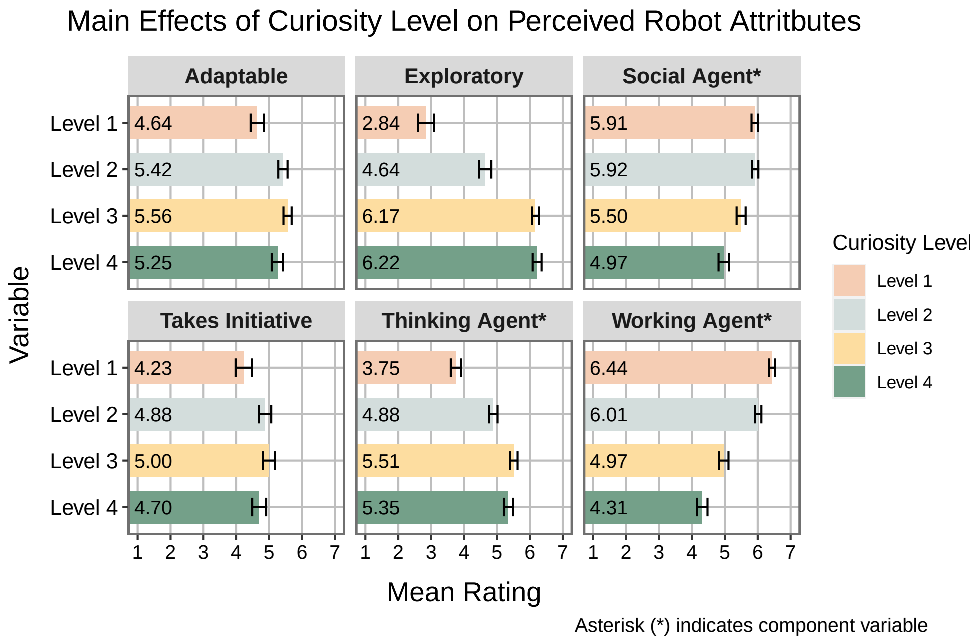


Figure 4.2: Impact of curiosity behavior on rating the robot as adaptable, exploratory, and taking initiative as well as a social agent, thinking agent, and working agent,.

we found significant difference between levels 1 vs. 2, $F(1{,}61){=}17.05$, $p{=}.001$, $\eta_p^2{=}.22$, and levels 3 vs. 4, $F(1{,}61){=}4.53$, $p{<}.05$, $\eta_p^2{=}.07$, but not between levels 2 vs. 3. We also captured a significant main effect of the robots' curiosity behaviors, $F(3{,}183){=}82.17$, $p{<}.001$, $\eta_p^2{=}.57$, on ratings of robot exploration. Here, contrasts revealed significant difference between levels 1 vs. 2, $F(1{,}61){=}50.25$, $p{<}.001$, $\eta_p^2{=}.45$, and levels 2 vs. 3, $F(1{,}61){=}52.76$, $p{=}.001$, $\eta_p^2{=}.46$, but not between levels 3 vs. 4. We found a significant main effect of the robots' curiosity behaviors, $F(3{,}183){=}3.25$, $p{<}.05$, $\eta_p^2{=}.05$, on ratings of the robot being likely to take initiative on new tasks. Here we only observed a significant difference between levels 1 vs. 2, $F(1{,}61){=}5.98$, $p{<}.05$, $\eta_p^2{=}.09$. There was no main effect for robot cognitive state or interaction effect observed relating to any of the aforementioned individual survey items. There was no effects between our independent variables and ratings of the robot as autonomous.

4.3 Qualitative Data Analysis

Below we present the process and results from our sentiment analysis and thematic coding of participants' open-ended responses describing the robots they saw and sharing what they liked/disliked about them. Our framing and understanding of the qualitative data in this current study stems directly from our insights in the previous study. As such, we will also refer back to our earlier reflections as well as provide some context. However, if additional details are needed please see section 3.3 for methods and 3.3.2.1 for findings.

4.3.1 Sentiment Analysis

As in the previous study, we began by generating emotional valence scores for participants' open-ended responses with VADER [25]. Next, we ran a mixed ANOVA using level of curiosity as our within-subjects independent variable and priming expectations of a robot's cognitive state (Curious vs. Learning vs. Autonomous) as our between-subjects independent variable to predict the emotional tone of participants' written responses. We captured a significant main effect a robot's curiosity, $F(3,183)=9.38$, $p<.001$, $\eta_p^2=.13$, on the emotional tone of participant responses. There was no main effect for a robot's cognitive state or interaction effect observed. We also ran repeated contrasts on each of the four levels of curiosity behaviors. Here we found significant difference between levels 3 vs. 4, $F(1,61)=8.19$, $p<.01$, $\eta_p^2=.12$, but none

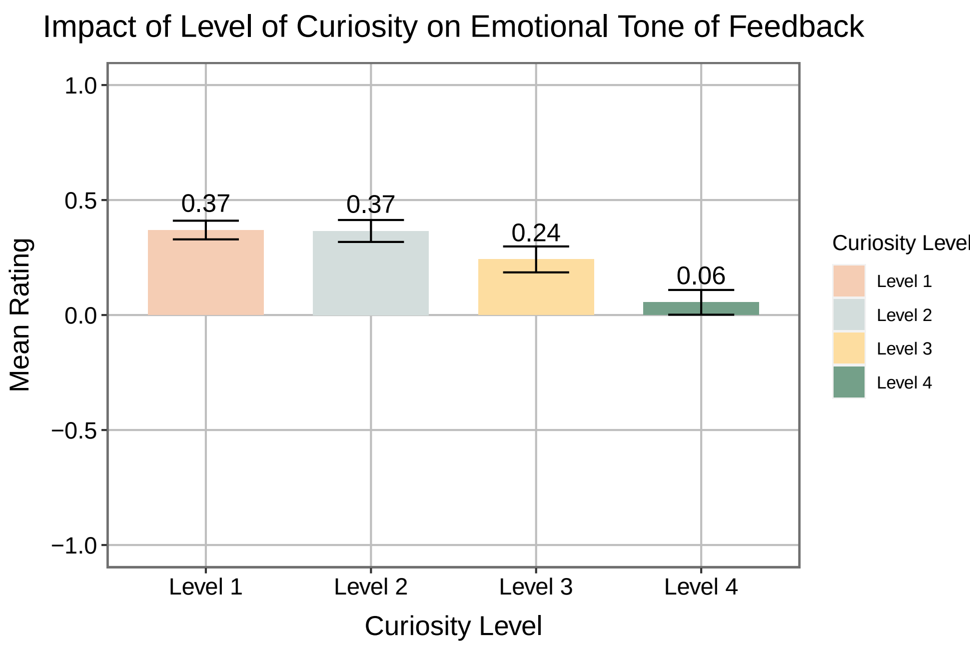


Figure 4.3: Robot curiosity behavior impacted emotional tone of participant open-ended responses. Significant change occurred at the highest level of curiosity.

between levels 1 vs.2 or 2 vs. 3. See Figure 4.3.

4.3.2 Thematic Coding

In Study 1, sentiment analysis revealed a significant change in the tone of participants response to the robots. Additionally, when we examined participants' open-ended responses, we saw a strong tendency to focus on assessments of a robot's obedience to its human counterpart and the mental states determining that relationship. While spending time exploring our Study 2 qualitative data, we found that participants, again, took a similar stance. However, following a grounded theory approach, informed by our Study 1 data, we were now able to see a subtler, and more complex, interplay between these two facets [7]. We recorded our codes and iterated until we reached satu-

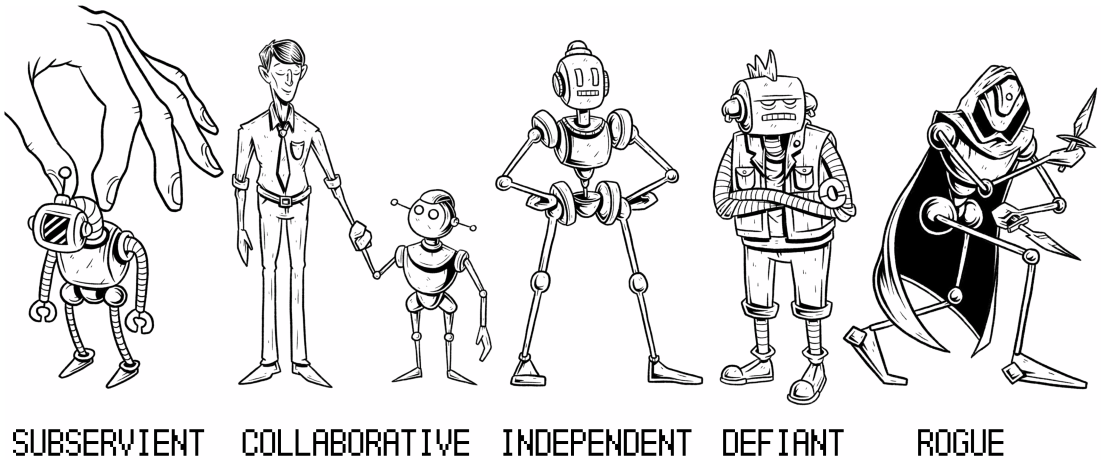


Figure 4.4: Thematic codes of robot-human dynamics in relation to the perceived levels of the robot's obedience and cognitive ability in participants' open responses.
Illustration by: Louis Riddick

ration (i.e., subsequent passes ceased to reveal new themes). From there we iterated our code-book, modifying meaning, identifying relationships between themes, and removing semantic redundancies until we reached consensus. Descriptions of meaning and examples were assigned to each code. Our final code-book consisted of five parallel-linked ordinal themes representing the range in which participants described a robot's action, cognition, and motivation as well as how these facilitate the perceived relationship to its human counterpart. These, in order from high obedience/low cognition to low obedience/higher cognition, were: Subservient, Collaborative, Independent, Defiant, and Rogue.

Next, we tested inter-rater reliability of our code-book. Participant ID, curiosity level, and priming conditions were removed and the data was presented as a randomized list. The primary coder blind-coded the full data set. Next, a second coder was trained on the code-book and blind-coded a random subset of 30% of responses (n = 77). Cohen's κ was found to be sufficiently high after the first pass (κ = .917) and no further iterations were deemed necessary. Coders had been instructed to apply a single code, from the set of five, to each participant response but in the event that the content was not applicable to simply label it as such (i.e., N/A). Upon closer inspection of N/A coded cells, we found that on any participant response who was tagged with N/ it was applied to more than of their responses. All of those participants' (n=*6*) responses (n=*24*) were excluded from final thematic analysis due to displaying evidence of low effort responses (e.g., simple non-varying or nonsensical responses) leaving us with a final coded data set of participants (n=*58*) and comments (n=*232*). Our final code book (Appendix E.1) consisted of five mutually exclusive codes. See Figure 4.4.

Subservient was used when participants described the robot as generally performing its task. While the robot may be seen as possessing some autonomy in its action or ability they ultimately viewed that autonomy as subservient to the will of its operator. This didn't necessarily indicate a view of the robot as entirely mindless, or even without agency, but rather that the human instructions supersede other underlying motivations the robot may have.

> It didn't pay any extra attention to see what's in the box and why it's there. It didn't care... The robot feels more like a tool or equipment and lost some of the human-like quality. It's very efficient but not very charming or interesting. –P13 (Curious prime/L1)

We saw that here participants still attributed mental capacities to the robot, such as attention, focus, and interest, but not in a way that didn't serve its immediate task. In these cases the participants often described the robot not as choosing to perform a task but as being compelled to, in some sense.

> The robot just followed orders and didn't worry about anything else... The robot probably wouldn't learn anything new and it doesn't have any independence. I would be more interested in a robot that is more adaptable. –P43 (Autonomous prime/L1)

Collaborative was used to capture a step above subservience in terms of self-will and mental capability while still being primarily attentive to the whims of the human. The collaborative robot was distinguished by the attribution of higher mental states coupled with frequent perception of choice. In other words, the robot, though primarily functioning to complete its task, made decisions not to be distracted by the box for the sake of prioritizing its assigned work. It may have been described as having a desire to learn and even deviating slightly to scratch that itch but ultimately putting the human need first.

> It stalled for a moment to look at something new but then went out to complete the task. I liked that it only took it a moment to look at the object and move on. It knew that it had a task at hand that it had to take care of and kept on going. –P7 (Curious prime/L2)

Independent could be seen, in some sense, as the middle point between the poles of obedience. Here participants described the robot as though it had free will to perform tasks. The robot was described in terms such as "human" or "independent." This represented a balanced point where the robot sought out and fulfilled its own curious interests while also seeing its responsibilities through.

> The robot seemed like a person because it looked at the box for a moment but didn't stop to analyze it. I like that it noticed the box but didn't stop because that's probably what a person would do, so as to not slow down the task they're trying to finish. –P56 (Autonomous prime/L2)

Moreover, these comments had a neutrality, or observational tone, about them. In particular, the framing of the human moved from being master or controller to a separate entity who is soliciting the robot for assistance, if referenced at all. So the robot is viewed as being aware of its role in a group dynamic but also willing and capable of making time for itself, so to speak. Much in the way that people are clear individuals and capable of serving a collective interest.

> The robot is interested in what is going on around them, but also wants to complete the tasks in the office. I liked that the robot took time to examine other things in the office before completing the task. It made the robot seem more human-like. –P31 (Learning prime/L3)

Defiant moves into the territory of annoyance at the robot's willfulness and diminished cognitive state though it is still perceived to serve human interests. Here the robot is taking direction but no longer performing its task in an adequate fashion. Moreover, the perceived cause of this degraded performance coincides with a shift in the framing of a robot's mental attributes. As in our first study, we saw an increase in negatively framed comments about the robots' cognition coupled with a decrease in positive ones. Here we begin to see participants move from framing the robots' mind as thinking, interested, human-like, etc., to wandering, confused, child-like, or even dumb and ditsy. Though they may comment on the robot eventually completing its task the process causes friction because it is seen as prioritizing the human instruction too low.

> It is hesitant and curious. Maybe a bit distracted or confused. Seemed unsure of what it wanted to do or got distracted by the box. Took too long to complete the task because it was distracted. –P30 (Learning prime/L3)

> Bordering on incompetent and not understanding basic things. I didn't like how it went around a couple of times to look at the box. It made it seem dumb. –P3 (Curious prime/L4)

Rogue was used to describe the final stage of perceived robot disobedience.

In these instances the robot had crossed beyond defiance and was seen as not even considering the human instructions. In other words, the robot is perceived as fully self-directed and no longer compelled by subservience to the human counterpart. Though it finishes its assigned task participants here viewed it as an afterthought in such a way that the robot had disregarded the importance of its work or that it chose not to care.

> It responded to the request but took it's time getting there. Now I am wondering why it took so much time to look at a box on the floor. If there were people standing around would it take time to investigate them too? What a waste of time. It knows where to go and should just go there. –P29 (Learning prime/L4)

These comments were also peppered with reflection on the robots' cognition but, as in our first study, now solidified in a negative light. Interestingly though, the negative takes on the robots' cognition here weren't reflected as another decrease in intellect, as in the previous category.

> It is inquisitive and irresponsible. I didn't like that it wasn't doing what it was meant to do. –P23 (Learning prime/L4)

In fact, what we saw was as an increase in perceived mental ability, compared to the Defiant robot. However, despite its intellect its efforts ceased to hold any tangible value in the eyes of participants and even prompted concern or suspicion of the robots' motives. That, coupled with the aforementioned views of extreme willfulness and disobedience, led to some of the more glowering comments.

> This robot was overly curious about its surroundings. It is completely willing to delay completing its mission in order to fulfill its own curiosity. I did not like that the robot had the audacity to explore the box for such a long time. It wasted the letter writer's time and accomplished nothing for itself. –P32 (Learning prime/L4)

We ran a χ^2 analysis to assess whether the frequency of each code differed across each level of robot curiosity behavior. Across the four levels, we found statistically significant differences in the number of people who viewed the robot as; subservient (χ^2=65.13, p=.0000), collaborative (χ^2=18.71, p=.0003), independent (χ^2=15.27, p= .0016), defiant (χ^2=37.62, p =.0000), and rogue (χ^2=25.33, p=.0000).

Next, to see where significant change occurred, we ran pair-wise χ^2 tests. We found significant differences in the number of people who saw the robot as being; subservient between levels 1 vs. 2, (χ^2=10.59, p=.0011) and levels 2 vs. 3 (χ^2=11.84, p=.0006), collaborative between levels 1 vs. 2 (χ^2=5.00, p=.0253) and 3 vs. 4 (χ^2=7.35, p=.0067), independent levels 2 vs. 3 (χ^2=4.80, p=.0285), defiant levels 2 vs. 3 (χ^2=10.29, p=.0013), and rogue levels 3 vs. 4 (χ^2=4.55, p=.0330). Our χ^2 tests did not reveal a significant difference for any of the thematic codes across cognitive primes.

For frequency of thematic codes in each between-subject priming condition across curiosity behavior levels see: Figure 4.5.

4.4 Summary of Study 2 Findings

Study 2 provided further evidence that intentionally designed robot behaviors can appropriately convey analogous internal states. Here we tested hypotheses about how well increasingly curious robot behaviors would match people's expectations when they were primed to anticipate a curious robot, a learning robot, or an autonomous robot. We found support for H2, when primed for a curious robot people's expectations

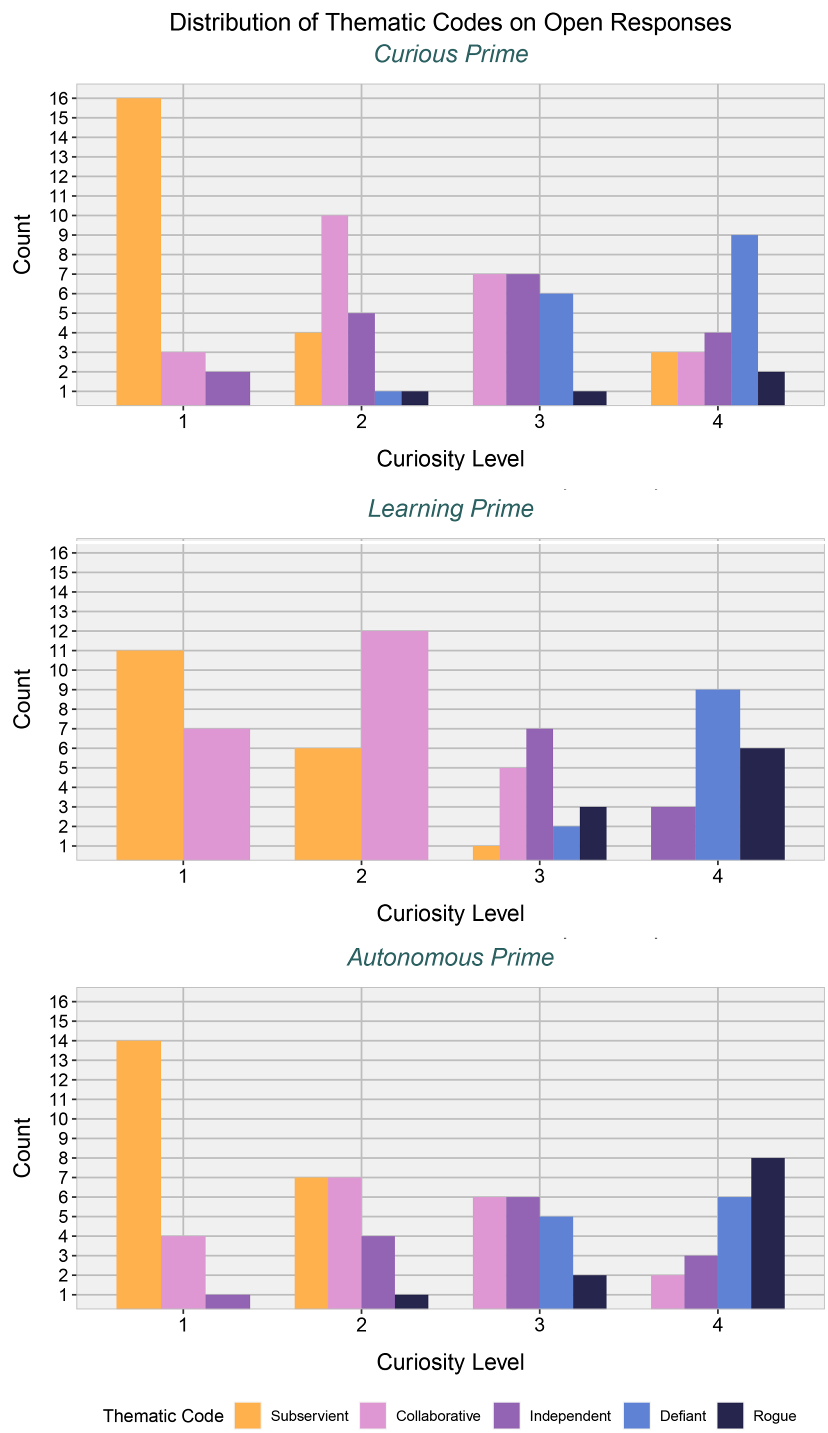


Figure 4.5: Frequency of thematic codes across curiosity behavior levels in (top to bottom): Curious, learning, and autonomous prime conditions.

were more effectively matched as the level of curious behavior went up. We also found support for H3, when primed for a learning robot people's expectations were most effectively matched in the low level of curiosity (Level 2) and then began to fall off at each subsequent level. We also found support for H4, when primed for an autonomous robot participant expectations would be most effectively matched when the robot showed no or little curiosity. Overall though in support of H1, people were found to hold unique expectations of external behaviors in the same robot based solely on single word primes about its internal mental state.

Unfortunately, in Study 2, priming participants to expect specific cognitive qualities from a robot did not offset negative changes to general perceptions of the robot (RQ1) as a social or thinking agent. Indeed, we observed near identical downward trends in ratings of the robot as both a social (likability in Study 1) and working (competent, responsible, focused in Study 1) agent while ratings of its thinking ability (curiosity, drive-to-learn in Study 1) went up, regardless of priming conditions (See: Figure 3.2 and Figure 4.2). Our analysis of sentiment scores in open ended response questions, again, paralleled the survey data and we saw an overall negative trend in the emotional tone of their feedback as curiosity increased. See: Figure 4.3. However, priming people to anticipate a robot possessing specific cognitive profiles (autonomous vs. learning vs. curious) and the level of curious behavior did significantly impact how well the robots' behaviors met their expectations (See: Figure 4.1).

Our examination of open-ended responses, informed by insights gleaned in Study 1, highlighted a shift in the perceived robot-human relationship as curiosity increased. Our thematic analysis revealed a conceptual scale of five parallel-linked codes (See: Figure 4.4) describing the way participants view the robot in relation to its human counterpart changed across levels of curiosity largely in relation to its obedience and cognitive ability, as in Study 1. We observed that, in the participants' eyes, as the robot gained increased mental capacity it also shifted away from a dynamic in which the it needed to work wholly in service to human whims. Interestingly, this was not a purely linear change wherein the robot exhibited more curiosity, seemed smarter, and also seemed less obedient. See Figure 4.6.

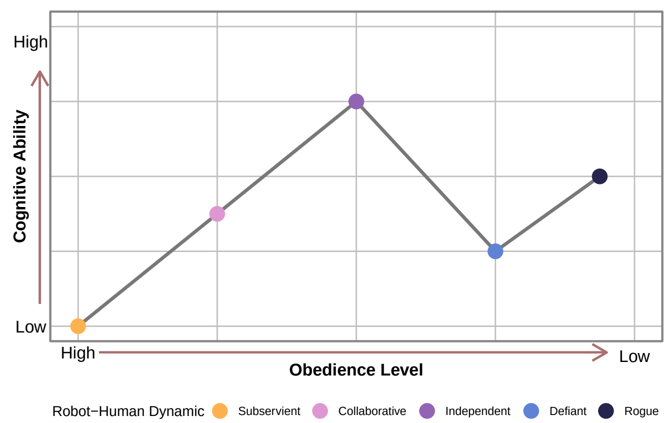


Figure 4.6: Changes in conceptual themes of robot-human relational dynamics along axes of cognitive ability and obedience (low to high).

Rather what we saw was that in the mid levels of curious behavior the robot seemed to reach a peak of both autonomy and mental ability while still being seen as competent to function within a group dynamic, much in the same way a human being is an individual who functions in a group. However, after that point a robot's perceived intelligence became tainted first with a view of it having a reduced mental ability (e.g., dumber), before turning upwards–figuratively–again and taking on a deviant quality.

Chapter 5

Where the Thesis Ends

In this thesis, we presented findings from two experiments, showing that there may indeed be such a thing as too much curiosity in robots. As a social agent, our robot became significantly less likeable at the highest level of curious expression, in both studies. Similarly, it was perceived as being less competent, as a working agent. This was especially true when it was supposed to be on the job in Study 1 (see Figure 3.3). This suggests that curious robot exploration will be more socially acceptable if the robot does that exploration while off-duty and not assigned to a particular work task. Below we discuss our key findings from both studies as well as some insights gleaned in designing the curious robot behaviors in collaboration with a professional animator. We will then end with a reflection on some of the limitations of both studies as well as possibilities for future work.

5.1 Conclusion

As we develop curiosity in robotic systems, it will be important to take the perspectives of coworkers and bystanders into account. If these robots are going to succeed in performing their tasks, especially their learning tasks, they may need instruction, guidance, and feedback from the people in their work environments. In particular there is a growing call for HRI research to look for ways to create harmonious interactions wherein both the robot and human agents' needs are met, as opposed to a paradigm where a robot achieving its goal without overtly bothering the human is considered successful [55].

Fortunately, there are creative ways to approach this from both an algorithmic and interaction design space. For instance, in a recent study, a robotic agent–in an online virtual setting–incorporated a model of individual human helpfulness to outperform other, traditional, ML approaches by asking human counterparts for help less often but at more optimal times [42]. The same study found that people had more negative social reactions to the robots that didn't adapt to individual and contextual factors when asking for help [42]. It seems that before we try and have lay-people offer guidance to robots, we may need to better understand how expressing internal mechanisms of learning (e.g., curiosity) is perceived by human stakeholders–both direct and indirect. The current work offers insight into how expressing curiosity will affect people's perceptions of robots in work contexts, including their feelings toward them.

Digging deeper, in Study 1 and 2 there were often significantly negative differences between medium curiosity (L3) and high curiosity (L4). However, there were often no significantly negative differences between no curiosity (L1) and low curiosity (L2). Since there was a significant increase in the perceived cognitive ability of the robot at each stage, it may be that more subtle forms of curious expression will be more socially acceptable.

While we demonstrated that people attributed increasing levels of cognition to the robot as it behaved with more curiosity, in Study 1, we also questioned whether the behaviors we designed to register as increasing robot curiosity truly equated to curiosity in people's minds or if the exploratory behaviors we designed would register as appropriate when given a similar self-driven mental capability label. In other words, were these behaviors truly seen as curiosity or would they match any human-learning model buzzword (e.g., learning, autonomous, smart)? Further, we were interested in whether expectation setting could act as a social intervention to offset the previously captured decline in perceived likability and competence when the robot was on-duty.

In Study 2 we found that when a person was primed to anticipate a curious robot, the more curious behavior it exhibited the more the robot matched their expectations. An autonomous robot, on the other hand, was most matched when the robot exhibited no or very little curious behavior and quickly dropped off after that. Since learning and curiosity are often treated as interchangeable adjectives, we were particularly interested in whether the learning condition would follow the same trajectory as the curious robot (i.e., as curious behavior increased so did expectation matching). We found that expectations of a learning robot followed its own unique trajectory peaking early and then steadily dropping off as the robot exhibited increasing amounts of curious behavior.

This suggests that when designing the internal cognitive models of a robot (i.e., how the robot thinks/learns), the relevant external qualities (i.e., how the robot acts/moves) should be simultaneously considered with the same level of thoughtfulness. In the future, when artificial learning processes for a robot are based on specific models of human cognition then the robot should be carefully outfitted with the ability to express appropriate behavioral analogs. This may allow human understanding of, and interaction with, a robot to be intuitively informed. Furthermore, making sure to match human expectation with the robot's actual behaviors and abilities can help prevent expectation violations on the part of the human. What we captured in these studies suggests that alignment of both mind and body is necessary for a harmonious robot.

An unexpected insight from Study 1 was the spontaneous mentions of more curious robots being more disobedient and having notable cognitive traits. Robot disobedience and apparent cognitive abilities were not topics that we had set out to examine, but they merit exploration in future work. Guided by our Study 1 qualitative analysis, we found a similar trend in our Study 2 open-ended responses and developed a conceptual framework of five parallel-linked ordinal themes explaining how increasing cognitive ability coupled with decreasing obedience affects the perceived relationship of the robot to the human. These encapsulated the shift of the robot, in the participants' eyes, from an obedient robotic tool with limited–but some–cognition to a balanced individual non-human agent with high cognition and the ability to function in a group setting. From then a robot's free-will became a liability, reducing perceptions of its cognitive ability (e.g., seeming dumb) and making it seem irresponsible, before transitioning to be viewed as willfully disregarding the human counterpart but also with an increased mental ability, albeit a devious one. Furthermore, the trends for these codes differed depending on what kind of robot participants were expecting (See Figure 4.5).

This empirical work sheds light onto how different levels of curious robots' behaviors will affect people's perceptions of robots in the workplace. These results can be used to inform the design of robots that engage in curiosity-driven learning. In optimizing for lifelong learning, it will be important to include robot learning goals as well as considerations of how much curious exploration will be acceptable by the people in the robot's environment.

5.2 Designing Alignment of Body and Mind

While character animation techniques have long been used in the animation of 3-D computer graphics [34], the translation of animation techniques to robot motion design is less straightforward. Unlike on-screen characters, robots do not have infinite degrees of freedom and real world robots need to abide by the laws of physics. While the HRI community is making headway into programmatically producing more human-readable motion trajectories [8,20,24,33], there are still many challenges ahead as robots not only work with human collaborators, but also encounter bystanders in the workplace, who are not necessarily directly interacting with them (e.g., staff who do not directly benefit from using the robot). These robots need to be socially acceptable by everyone the workplace, including those indirect stakeholders [10, 11], if they are going to stand a chance at being adopted for long-term use.

There are some formalizable guidelines for curious robot behavior design, like using gaze at a target to indicate minimal curiosity, adding body rotation toward the target to indicate more curiosity, and adding full body pausing and circles around a target to indicate even more curiosity. However, there are other animation techniques employed that might be harder to formalize with existing approaches to motion planning for robot expressivity (e.g., deciding how to glance vs. look vs. stare). At least two approaches to incorporating those other techniques include: (1) bringing character animators onto robot motion design teams as full collaborators and (2) gathering enough examples of robot motions designed by animators to train machine learning algorithms to produce more clearly expressive robot motion trajectories.

5.3 Limitations & Future Work

A limitation of the studies we presented here is that they were run online instead of in-person so we do not yet know exactly how these findings will generalize to real world robot deployments. Prior HRI research comparing in-person to on-screen robotic agents has shown that the direction of effects is often similar between the two (e.g., [50,70]), but effects can vary by robot interaction scenario type (e.g., task-oriented vs. conversation-oriented [23]. Furthermore, people tend to be more engaged with [1,50] and feel more empathy for [56] a collocated physical robot than for an on-screen one. As these studies were conducted between Summer 2020 and Spring 2021, we were limited in our ability to run in-person studies due to the COVID-19 pandemic. We chose to use an on-screen prototype to study these curious robot movements (as recommended for movement-centric robot designs [21]. We made sure to produce behaviors that worked within the degrees of freedom and joint limits of the real Kuri hardware platform so that it will be possible for us to run follow-up studies, using the same curious behaviors, using the real world Kuri robots.

Another limitation is that the robot was only shown performing one specific task in one setting. It is entirely conceivable that people don't want a delivery robot to be curious while it is working but that in other contexts it might be more acceptable. Additionally, our displayed curiosity was–by design–mundane and without clear benefit. What if the curious exploration had yielded some tangible benefit (e.g., notice and address a safety hazard) for the robot and/or its human counterpart? Further work would benefit from exploring other settings, tasks, and outcomes for a curious robot.

Similarly, we only tried one intervention (i.e., priming expectations of the robot) in an attempt to offset negative views of robot curiosity. Future work should examine other social interventions such as presenting the robot as on its first day on the job. Would observers be more generous when evaluating curiosity at a time when it is beneficial to be especially inquisitive?

Participants in our experiments also viewed the robot from a bystander perspective (i.e., third person perspective) rather than from a direct interactant perspective.

While this could be a limitation, it is actually the perspective that most people will have of these robots and indirect stakeholders (e.g., bystanders) [11] are often important for considering in the design of real interactive systems.

As we mentioned in section 3.1, we were unable to hold constant the time duration of each video. The robot simply needed more time to display its curiosity at the higher levels. In real world deployments, robots that do more curious exploration will likely take more time to explore than robots that do less (or no) curious exploration.

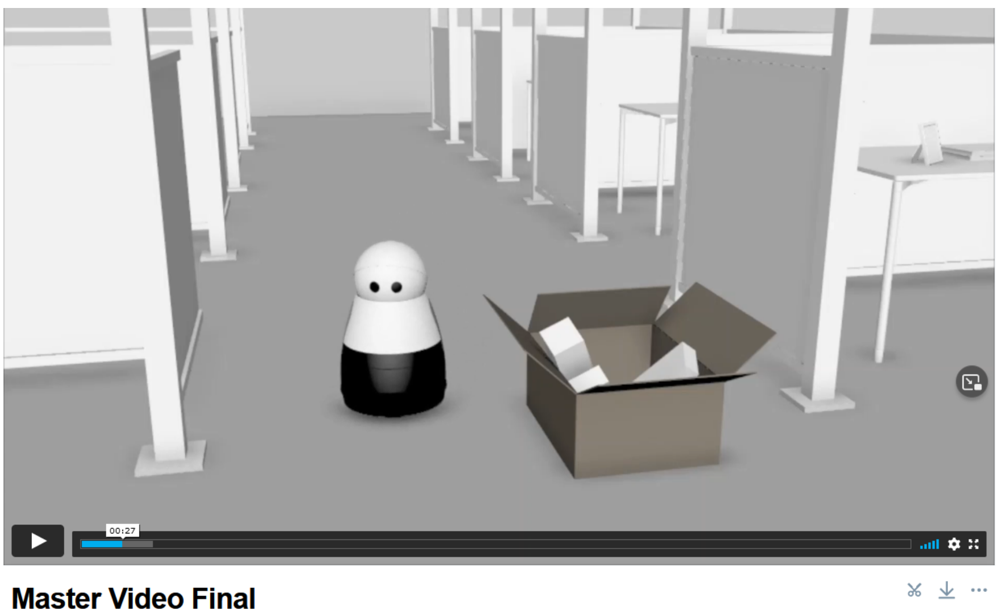
As with many HRI studies, these experiments only used one type of robot in one type of setting with only English-speaking participants. To assess the generalizability of these results, it will be important to replicate this study with other hardware platforms (e.g., robot arms) in other settings (e.g., classrooms, homes, hospitals) and with broader user groups (e.g., with other cultural groups). One of the strengths of these online studies is that they are more readily replicable. We have shared our study materials so that anyone who is interested in running their own experiments with a similar framework can use the stimuli provided in Appendix A.

The short-term nature of this study may make the results subject to novelty effect biases. We are not yet at the stage of having enough confidence in the design of curious robots to know if the long-term deployments will be pursuing in its current form. If we had found extremely positive responses to the extremely curious robot behaviors, we would have been skeptical of these results simply showing how engaging a brief animation of a robot could be. However, most of these findings showed negative reactions to the curious robot behaviors. It is possible that people might come to see the benefit of curious robots over time as their performance improves or they become more endearing over time. As such, we are currently planning longer-term, real world deployments on this hardware platform.

For future work, we will be exploring more behavioral types of effects, not only perceptions and self-reported experiences. In our ongoing collaborations, we currently have plans to deploy these robots in real office settings. Once it is safe for people to collocate in offices again, we will be able to run in-person studies to examine behavioral responses such as willingness to help and teach these curious robots.

Appendix A

Video Stimuli



Click on the image above to view video stimuli on Vimeo. Videos were created in collaboration with Doug Dooley, a character animator specializing in design of robot behaviors.

Appendix B

Study 1 Tables

Curiosity Level	**Level 1**		**Level 2**		**Level 3**		**Level 4**	
On vs. Off duty	**On**	**Off**	**On**	**Off**	**On**	**Off**	**On**	**Off**
Curiosity***	3.94 (0.55)	2.92 (0.58)	4.47 (0.34)	4.62 (0.51)	5.82 (0.39)	6.23 (0.26)	5.47 (0.37)	5.92 (0.46)
Likability**	5.82 (0.30)	6.04 (0.16)	6.01 (0.20)	6.02 (0.21)	5.47 (0.38)	5.42 (0.30)	3.94 (0.29)	5.12 (0.32)
Drive to learn**	3.78 (0.38)	3.10 (0.55)	4.61 (0.29)	4.03 (0.45)	5.78 (0.23)	5.54 (0.31)	5.16 (0.21)	5.49 (0.43)
Self-sufficiency**	4.69 (0.36)	4.51 (0.46)	5.06 (0.28)	4.44 (0.49)	5.00 (0.35)	5.23 (0.45)	4.00 (0.23)	5.03 (0.40)
Focus*	6.29 (0.25)	5.92 (0.56)	5.12 (0.44)	5.46 (0.35)	3.65 (0.44)	3.46 (0.61)	1.82 (0.35)	0.35 (0.70)
Competence*	5.88 (0.36)	5.31 (0.57)	6.18 (0.26)	5.54 (0.51)	5.12 (0.40)	4.69 (0.56)	3.29 (0.29)	4.46 (0.55)
Responsibleness*	6.35 (0.24)	5.92 (0.38)	6.06 (0.22)	5.85 (0.25)	5.41 (0.33)	5.23 (0.34)	3.76 (0.36)	4.46 (0.43)

Table B.1: Study 1 Descriptive statistics: Mean and (Standard Error) values for each cell in the experiment design.

Key: * = individual item, ** = component, *** = Manipulation check.

Component	**Drive to Learn**	**Likability**	**Self-Sufficiency**
Items (In order)	Explores Surroundings Desires New Knowledge Learns On Own	Acceptable Likable Trustworthy Approachable	Initiative Independent Intelligent
Loading Level 1	0.92, 0.95, 0.93	0.81, 0.73, 0.87, 0.85	0.88, 0.83, 0.71
Loading Level 2	0.94, 0.92, 0.76	0.84, 0.84, 0.87, 0.78	0.91, 0.83, 0.88
Loading Level 3	0.81, 0.85, 0.77	0.77, 0.86, 0.85, 0.87	0.88, 0.91, 0.88
Loading Level 4	0.75, 0.9, 0.75	0.86, 0.83, 0.82, 0.77	0.82, 0.79, 0.88
Reliability	$\alpha = 0.92, 0.85, 0.72, 0.71$	$\alpha = 0.82, 0.85, 0.84, 0.83$	$\alpha = 0.72, 0.83, 0.87, 0.76$

Table B.2: Study 1 Principal Component Analysis.

Loading scores for each set of items and reliability at each level of robot curiosity

Appendix C

Study 1 Code-Book

Code	Tag	Explanation of Code	Examples
Focused	**FOC**	Definition: Discussion of HOW Kuri arrives back to the charger or to the point of completing its task When to use: Participant describes Kuri as focused, on-task and/or efficient while on task	"It got straight to the task and wasn't distracted by anything."
Unfocused	**UNFOC**	When to use: Participant describes Kuri as in a state of distraction, unfocused, etc., while on task or otherwise	"It's overly distracted and wastes too much time."
Reliable	**REL**	Definition: Discussion on Kuri's ability to successfully perform tasks now or in the future When to use: Participant states or implies that Kuri can be depended to complete tasks	"Quick to perform tasks and would be able to overcome obstacles in the future."
Unreliable	**UNREL**	When to use: Participant states or implies that Kuri is failing to complete tasks and/or will fail to complete tasks	"Faulty and undependable. Untrustworthy to do what is needed."
Obedient	**OBE**	Definition: Discussion on Kuri's ability or willingness to listen to commands and instructions When to use: Participant states that Kuri is doing what it has been asked to do by the user or programmed to do	"Obedient and does what it is programed to do."
Disobedient	**DISOBE**	When to use: Participant states that Kuri is doing something other than what it should do and implies this is due to a character flaw, malfunction or choice	"Incompetent in following directions."
Positive Cognition	**SMRT**	Definition: Discussion of Kuri's cognitive abilities and/or traits When to use: Partipant applies positive/useful/higher cognitive traits to Kuri	"Learning, the robot learns from its environment to make future trips quicker."
Negative Cognition	**STUPID**	When to use: Participant applies negative/unproductive/lesser cognitive traits to Kuri	"Extremely slow and easily distracted. I would call it 'ditzy."'
No Code	**NONE**	When to use: The response did not include any of these codes	"Calm and casual."

Table C.1: Study 1 Thematic Code-book

Appendix D

Study 2 Tables

Component	**Thinking Agent**	**Working Agent**	**Social Agent**
Items (In order)	Explores Surroundings, Desires New Knowledge, Learns On Own, Independent, Curious	Focused*, Competent*, Obedient*, Unobtrusive*, Acceptable, Follows Instructions	Trustworthy, Likable, Responsible, Approachable
Loading Level 1	0.89, 0.91, 0.71, 0.57, 0.86	0.76, 0.74, 0.87, 0.80, 0.75, 0.80	0.81, 0.73, 0.72, 0.69
Loading Level 2	0.83, 0.85, 0.61, 0.66, 0.76	0.64, 0.71, 0.80, 0.66, 0.74, 0.78	0.81, 0.85, 0.84, 0.76
Loading Level 3	0.68, 0.88, 0.52, 0.71, 0.83	0.55, 0.81, 0.88, 0.65, 0.86, 0.90	0.94, 0.90, 0.91, 0.88
Loading Level 4	0.50, 0.86, 0.68, 0.79, 9.86	0.64, 0.81, 0.93, 0.51, 0.83, 0.88	0.91, 0.89, 0.91, 0.74
Reliability	$\alpha = 0.85, 0.80, 0.77, 0.79$	$\alpha = 0.82, 0.78, 0.86, 0.85$	$\alpha = 0.70, 0.83, 0.92, 0.89$

Table D.1: Study 2 Principal Component Analysis.

Appendix E

Study 2 Code-Book

Name	Tag	Description	Example
Subservient	SUB	Robot performs all tasks assigned. While it may possess autonomy in action its autonomy is subservient to the will of its operator. This is the master slave model. The robot doesn't necessarily choose to perform its task it is compelled.	The robot's behavior what what I would expect. It was summoned to go to a specific room and fetch mail. It immediately traveled to the destination and picked up the mail. I liked that it followed the instructions. Even when there was a distraction (a box with things in it) on the floor, it went right past it to the destination.
Collaborative	COLAB	Robot acts in conjunction with its operator or partner but it is still in service to the partners objectives. The robot is attentive to its own needs but not at the expense of the task at hand. The robot chooses to perform assigned tasks.	It stalled for a moment to look at something new but then went out to complete the task. I liked that it only took it a moment to look at the object and move on. It knew that it had a task at hand that it had to take care of and kept on doing
Independent	INDI	The robot is independent. Sometimes even described as being human like. Right in the middle of disobedient and obedient. It is seen as having free will to perform tasks. Taking action on the part of human is seen as a choice. The human role moves from being framed as an operator or commander to a separate entity who is soliciting the robot for help.	The robot is interested in what is going on around them, but also wants to complete the tasks in the office. I liked that the robot took time to examine other things in the office before completing the task. It made the robot seem more human-like.
Defiant	DEF	The robot takes direction but makes a choice not to perform in an adequate way. It may ultimately complete a task but causes friction by prioritizing the human instruction below its own desires.	The robot seemed very distracted by the box. It seemed more like a child than a robot. I didn't like how much time the robot spend looking at the box because it slowed down the task it was supposed to do.
Rogue	ROG	The robot is fully self directed. It seen as prioritizing its own drives with little or no consideration to the human instructions. May still finish task but it is seen as an afterthought in a way where the robot has disregarded its importance.	curious and nosy. I don't think he should have been making the human wait while it checked out the box

Table E.1: Study 2 Thematic Code-Book

Bibliography

[1] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[2] Liam J Bannon. From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In *Readings in human–computer interaction*, pages 205–214. Elsevier, 1995.

[3] D. E. Berlyne. Curiosity and exploration. *Science*, 153(3731):25–33, jul 1966.

[4] Daniel Ellis Berlyne. A theory of human curiosity. *British Journal of Psychology. General Section*, 45(3):180–191, 1954.

[5] Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law. Expression of curiosity in social robots: Design, perception, and effects on behaviour. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[6] Amy Chak. Teachers' and parents' conceptions of children's curiosity and exploration. *International Journal of Early Years Education*, 15(2):141–159, 2007.

[7] Kathy Charmaz. *Constructing grounded theory.* sage, 2014.

[8] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

[9] Sébastien Forestier, Yoan Mollard, and Pierre-Yves Oudeyer. Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*, 2017.

[10] Batya Friedman and David G Hendry. *Value sensitive design: Shaping technology with moral imagination.* MIT Press, 2019.

[11] Batya Friedman, Peter Kahn, and Alan Borning. Value sensitive design: Theory and methods. *University of Washington technical report*, (2-12), 2002.

[12] Yogesh Girdhar and Gregory Dudek. Modeling curiosity in a mobile robot for long-term autonomous exploration and monitoring. *Autonomous Robots*, 40(7):1267–1278, 2016.

[13] Goren Gordon. Social behaviour as an emergent property of embodied curiosity: a robotics perspective. *Philosophical Transactions of the Royal Society B*, 374(1771):20180029, 2019.

[14] Goren Gordon, Cynthia Breazeal, and Susan Engel. Can children catch curiosity from a social robot? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 91–98, 2015.

[15] Emily M Grossnickle. Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, 28(1):23–60, 2016.

[16] John Harris and Ehud Sharlin. Exploring the affect of abstract motion in social human-robot interaction. In *2011 Ro-Man*, pages 441–448. IEEE, 2011.

[17] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.

[18] Guy Hoffman. Dumb robots, smart phones: A case study of music listening companionship. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 358–363. IEEE, 2012.

[19] Guy Hoffman and Cynthia Breazeal. Cost-based anticipatory action selection for human–robot fluency. *IEEE transactions on robotics*, 23(5):952–961, 2007.

[20] Guy Hoffman and Cynthia Breazeal. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 1–8, 2007.

[21] Guy Hoffman and Wendy Ju. Designing robots with movement in mind. *Journal of Human-Robot Interaction*, 3(1):91–122, 2014.

[22] Guy Hoffman and Gil Weinberg. Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3097–3102. 2010.

[23] Laura Hoffmann and Nicole C Krämer. Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. *International Journal of Human-Computer Studies*, 71(7-8):763–774, 2013.

[24] Chien-Ming Huang and Bilge Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 83–90. IEEE, 2016.

[25] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.

[26] Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, Pierre-Yves Oudeyer, et al. Learning to recognize objects through curiosity-driven manipulation with the icub humanoid robot. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pages 1–8. IEEE, 2013.

[27] Jamie Jirout and David Klahr. Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental review*, 32(2):125–160, 2012.

[28] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[29] Frederic Kaplan. In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, 1(1):225–236, nov 2007.

[30] Todd B Kashdan, David J Disabato, Fallon R Goodman, and Patrick E McKnight. The five-dimensional curiosity scale revised (5dcr): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences*, 157:109836, 2020.

[31] Todd B Kashdan and Paul J Silvia. Curiosity and interest: The benefits of thriving on novelty and challenge. *Oxford handbook of positive psychology*, 2:367–374, 2009.

[32] Celeste Kidd and Benjamin Y. Hayden. The psychology and neuroscience of curiosity. *Neuron*, 88(3):449–460, nov 2015.

[33] Przemyslaw A Lasota and Julie A Shah. Analyzing the effects of human-aware motion planning on close-proximity human–robot collaboration. *Human factors*, 57(1):21–33, 2015.

[34] John Lasseter. Principles of traditional animation applied to 3d computer animation. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 35–44, 1987.

[35] Edith Law, Vicky Cai, Qi Feng Liu, Sajin Sasy, Joslin Goh, Alex Blidaru, and Dana

Kulić. A wizard-of-oz study of curiosity in human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 607–614. IEEE, 2017.

[36] Jordan Litman. Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition & emotion*, 19(6):793–814, 2005.

[37] George Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1):75–98, 1994.

[38] Wallace H Maw and A Jon Magoon. The curiosity dimension of fifth-grade children: A factorial discriminant analysis. *Child development*, pages 2023–2031, 1971.

[39] Wallace H Maw and Ethel W Maw. Establishing criterion groups for evaluating measures of curiosity. *The Journal of Experimental Education*, 29(3):299–305, 1961.

[40] Wallace H Maw and Ethel W Maw. Children's curiosity and parental attitudes. *Journal of Marriage and the Family*, pages 343–345, 1966.

[41] Bilge Mutlu and Jodi Forlizzi. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 287–294. IEEE, 2008.

[42] Amal Nanavati, Christoforos Mavrogiannis, Kevin Weatherwax, Leila Takayama, Maya Cakmak, and Siddhartha S Srinivasa. Modeling human helpfulness with individual and contextual factors for robot planning.

[43] Hung Ngo, Matthew Luciw, Alexander Forster, and Juergen Schmidhuber. Learning skills from play: artificial curiosity on a katana robot arm. In *The 2012 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2012.

[44] Behnaz Nojavanasghari, Tadas Baltrusaitis, Charles E Hughes, and Louis-Philippe Morency. The future belongs to the curious: Towards automatic understanding and recognition of curiosity in children. In *WOCCI*, pages 16–22, 2016.

[45] Pierre-Yves Oudeyer. Computational theories of curiosity-driven learning. *arXiv preprint arXiv:1802.10546*, 2018.

[46] Pierre Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 2009.

[47] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

[48] Hee Sun Park, René Dailey, and Daisy Lemus. The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4):562–577, 2002.

[49] Maddie Pelz and Celeste Kidd. The elaboration of exploratory play. *Philosophical Transactions of the Royal Society B*, 375(1803):20190503, 2020.

[50] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a com-

puter agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 145–152, 2007.

[51] Johnmarshall Reeve. The face of interest. *Motivation and Emotion*, 17(4):353–375, 1993.

[52] Johnmarshall Reeve and Glen Nix. Expressing intrinsic motivation through acts of exploration and facial displays of interest. *Motivation and Emotion*, 21(3):237–250, 1997.

[53] Pericle Salvini, Cecilia Laschi, and Paolo Dario. Design for acceptability: improving robots' coexistence in human society. *International journal of social robotics*, 2(4):451–460, 2010.

[54] Henk Schmidt and Jerome Rotgans. Epistemic curiosity and situational interest: Distant cousins or identical twins? *Educational Psychology Review*, 2020.

[55] Katie Seaborn, Norihisa P Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. Voice in human–agent interaction: A survey. *ACM Computing Surveys (CSUR)*, 54(4):1–43, 2021.

[56] Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. Poor thing! would you feel sorry for a simulated robot? a comparison of empathy toward a physical and a simulated robot. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 125–132. IEEE, 2015.

[57] Dajung Diane Shin and Sung-il Kim. Homo curious: Curious or interested? *Educational Psychology Review*, pages 1–22, 2019.

[58] Paul J Silvia. Interest—the curious emotion. *Current directions in psychological science*, 17(1):57–60, 2008.

[59] Paul J Silvia. Curiosity and motivation. *The Oxford handbook of human motivation*, pages 157–166, 2012.

[60] Satinder Singh, Andrew G. Barto, and Nuttapong Chentanez. Intrinsically motivated reinforcement learning. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04, page 1281–1288, Cambridge, MA, USA, 2004. MIT Press.

[61] Satinder Singh, Richard L Lewis, Andrew G Barto, and Jonathan Sorg. Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, 2(2):70–82, 2010.

[62] James P Stevens. *Applied multivariate statistics for the social sciences.* Routledge, 2012.

[63] Daniel Szafir, Bilge Mutlu, and Terrence Fong. Communicating directionality in flying robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 19–26. IEEE, 2015.

[64] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving

robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76, 2011.

[65] David R Thomas. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2):237–246, 2006.

[66] Sebastian Thrun and Tom M Mitchell. Lifelong robot learning. *Robotics and autonomous systems*, 15(1-2):25–46, 1995.

[67] Emre Ugur, Mehmet R Dogar, Maya Cakmak, and Erol Sahin. Curiosity-driven learning of traversability affordance on a mobile robot. In *2007 IEEE 6th International Conference on Development and Learning*, pages 13–18. IEEE, 2007.

[68] Gentiane Venture and Dana Kulić. Robot expressive motions: a survey of generation and evaluation methods. *ACM Transactions on Human-Robot Interaction (THRI)*, 8(4):1–17, 2019.

[69] Nick Walker, Kevin Weatherwax, Julian Allchin, Leila Takayama, and Maya Cakmak. Human perceptions of a curious robot that performs off-task actions. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 529–538, 2020.

[70] Sarah N Woods, Michael L Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials. In *ROMAN 2006-the 15th IEEE inter-*

national symposium on robot and human interactive communication, pages 51–58. IEEE, 2006.