Carnegie Mellon University

Characterizing the Role of Agent Identities in Interactions Among Individuals, Embodiments, and Services

Samantha Ida Reig

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Human-Computer Interaction Institute at Carnegie Mellon University

Committee:
Jodi Forlizzi (Co-Chair), Carnegie Mellon University
Aaron Steinfeld (Co-Chair), Carnegie Mellon University
Geoff Kaufman, Carnegie Mellon University
Terrence Fong, NASA Ames Research Center

CMU-HCII-23-101
May 2023

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This work was supported by a NASA Space Technology Research Fellowship (grant 80NSSC19K1133) and by the U.S. National Science Foundation (grants SES-1734456, IIS-2112633).

Keywords: interaction design, human-computer interaction, human-robot interaction, human-agent interaction, human-autonomy interaction, human-robot teaming, embodiment, re-embodiment, embodied agents, smart environments, smart homes, internet of things, user experience, personalization, mixed methods, quantitative, qualitative, user enactments, storyboards, social robots, service robots, conversational agents, virtual assistants, failure, trust, video study, story completion, loyalty, ancillary users, identity performance, mental models

Abstract

Samantha Ida Reig

Characterizing the Role of Agent Identities in Interactions Among Individuals, Embodiments, and Services

With the ongoing innovation of intelligent systems that coordinate and collaborate with humans, it becomes increasingly important to understand how interactions should be designed to support effective communication, social norms, and appropriately calibrated trust. These intelligent systems are becoming less and less constrained to single embodiments: voice-activated agents that are typically embodied in smart speakers, for example, can interact with users through multiple platforms and control multiple devices in a shared space. As researchers and designers explore the potential for agents to serve as interactive interfaces to complex systems, they grapple with questions of how technical constraints and social context might impact aspects of agents' design and use. These aspects include possibilities for and effects of physical design, how agents should handle complex ethical and interpersonal constructs like social privacy (what happens when a smart home agent keeps secrets?), how they might be mentally modeled (are they tools, collaborators, or something else?), and what their roles and responsibilities are among genuine social players.

I argue that agent identities can play a mediating role in shaping the interactions that are situated in these complex and integrated contexts, as well as their outcomes. By integrating theoretical, empirical, and design work on agent identity, smart environments, and technology mediation, I formulate a preliminary conceptual model of agent identity as a mediating entity in relationships among individuals, embodiments, and services. In this dissertation, I discuss several studies that explored possible future designs for agent identities as service touchpoints, manipulated agent identity in human-robot collaboration settings, examined the role of embodiment in interactions between agent identities and ancillary users, and explored possibilities for future human-agent interaction in smart homes and other smart environments. The work that I have done with my colleagues to date has revealed novel insights that aid in mapping the space of human-agent interactions in complex social and physical environments and informing new frameworks for understanding and studying AI agent identities. In interpreting these insights, we have made several contributions to basic scientific knowledge in human-computer interaction (HCI), human-robot interaction (HRI), and human-agent interaction (HAI). We also generated design implications for agents and robots that interact socially with people in various domains. This work has both revealed and addressed a crucial need and timely opportunity to formulate new, informed approaches to human-AI interactions that take a bird's-eye view of entire AI-rich environments (including multiple people, multiple agents, and multiple embodiments).

Acknowledgments

As they say, "It takes a village".

First and foremost, I want to thank my advisors, Jodi Forlizzi and Aaron Steinfeld, for their guidance and support, for keeping me on track, for demystifying the world of academia a little bit more with every advising meeting, and for reassuring me that things would work out. Jodi welcomed me to CMU before I even enrolled. In our time working together, she has given me so many of her hours and insights, supported my ideas and implored me to pursue them, and helped me put difficult things in perspective. Aaron welcomed me into the then-nameless Transportation, Bots, and Disability Lab when I first began my PhD, and it has been my "home base" on campus since. Through the years, Aaron has gone above and beyond in sharing his expertise, experiences, and restaurant recommendations; making himself available for last-minute meetings about paper-writing and career-navigating; and helping me network. In my mind, I compare Jodi and Aaron both to Glinda, the Good Witch of the North: along my path, there have been surprises that we've encountered together, but also truths that they've known all along and pushed (or graciously allowed, depending on the issue) me to discover for myself—and my journey through this PhD Oz has been all the better for it.

I am immensely grateful to my committee members, Terry Fong and Geoff Kaufman. Terry has been my fellowship mentor for many years. His support has been strong even from a distance, and he has consistently inspired me to reconsider the assumptions of my research and to think about human-robot interaction in new ways. Geoff's approaches to teaching, research, and community-building have had an immeasurable impact on my development as a student and a scholar of HCI, as an instructor, and as a person in the world. I thank them both for their feedback on this thesis.

Several other faculty members have influenced and inspired me during my time at Carnegie Mellon. My PhD was made possible in large part by John Zimmerman. In addition to bringing me onto his project and supporting all my work, he has shown me how to formulate innovative research questions, speak my mind in meetings, and use research-through-design methods effectively. I will always remember Chinmay Kulkarni's enthusiasm during our first conversation in his office as I considered the offer to join the program. Since then, his insights into my work (and our n=1 shared projects) have been invaluable. Raelin Musuraca and Motahhare Eslami helped me develop my teaching skills and cultivate my interest in teaching while modeling superb pedagogy. Jessica Hammer supported my research and service, and has always been a role model for me in so very many ways. Thank you all. Thank you also to the faculty who contributed to my work, helped me navigate the PhD program, and worked with me on department initiatives: Laura Dabbish, Jeff Bigham, David Lindlbauer, Nik Martelaro, Ken Holstein, Paul Pangaro, Stephanie Rosenthal, Jill Lehman, Niki Kittur, Scott Hudson, Henny Admoni, and Reid Simmons, among others.

Since pretty much day one, my go-to person for every question from "Should I worry about this confound?" to "Is the tone of this email appropriate?" has been Liz Carter. Liz is a powerhouse of a researcher and human. Having her as a mentor makes me feel like I won the jackpot. Thank you, Liz, for the pep talks, the paper edits, the reminders to go outside, and everything in between.

There is no possible way to say "thank you" enough to Queenie Kravitz. In the past six years, Queenie has handled countless departmental protocols and logistics. She has also encouraged me to trust what I knew, seek what I wanted, and prioritize what I needed. She is a hero of higher education, a brilliant collaborator, and a dear friend.

The depth and richness of the work described in Section II of this thesis was made possible by the contributions of four outstanding undergraduate students: Elsa Forberger, Danielle Oltman, Janet Wang, and Isabel Won, thank you for bringing your unique skillsets and perspectives to our summers together. I also thank Carnegie Mellon undergraduates and recent alumni Irene Kang, Winnie Lin, Benjamin Stone, and Thomas Von Davier for their work on preparing study materials and analyzing data.

I am grateful to my collaborators from beyond Carnegie Mellon. The experiences I have had working with Andrés Monroy-Hernández, Rajan Vaish, and Brian A. Smith have shown me how incredibly fulfilling good teamwork can be, and all three have provided unique perspectives on my dissertation that have helped me understand its place in HCI. Thank you to Tom Williams and Alexa Bejarano for identifying our mutual research interests and running with them.

Thank you to the mentors and collaborators at Cornell who sparked my interest in HCI and set me on a path to this degree. Mike Goldstein saw a sophomore psychology major unsure of how to handle a newfound love of coding and envisioned a human-robot interaction scholar. Thank you, Mike, for telling me that HRI was a field (and that it would welcome and nurture the merging of my interests) and for advising me on PhD applications. Thank you to Morten Christiansen for opening the door to research for me, and to Susan Fussell, Malte Jung, and Brett Stoll for bringing me into the fold and supervising and supporting my contributions to lab projects.

My colleagues in the PhD program, the TBD Lab, and the CMU HRI community are some of the coolest people in the world. I am so lucky to count Julia Cambre as my co-author, cohort buddy, office buddy, travel buddy, and, most importantly, friend. She has been my person through office after office, mini after mini, deadline after deadline, and thick and thin. This ride was made infinitely better and brighter thanks to her wisdom and support. Michal Luria's research approach and design skills have always inspired me; she is an amazing collaborator, and I am grateful for her leadership, advice, and friendship. I am grateful to my cohort-mates and friends Lynn Kirabo, Stephanie Valencia[2], Cori Faklaris, and Lea Albaugh for lending a listening ear on so many occasions, for providing feedback on dozens of practice talks, paper drafts, and study designs, and for enhancing my work with their keen insights at every turn.

I thank the newest (2022-2023) members of the "Craig St. Crew": Mai Lee Chang, Alicia Lee, Neeta Khanuja, and Hugo Simão. Collaborating with them—and with the AI-CARING team more broadly—over the past year has been intellectually stimulating, and the work so rewarding.

I want to thank the many other CMU friends and colleagues with whom I have shared offices, papers, organizational efforts, and stages at various points throughout these past six years: Alex Ahmed, Karan Ahuja, Sarthak Ahuja, Reuben Aronson, Abhijat Biswas, Sol Boucher, Vikram Kamath Cannanure, Daphne Chen, Tianying Chen, Prerna Chikersal, Judy Choi, Erica Principe Cruz, Anna Fang, Bailey Flanigan, Tesca Fitzgerald, Mary Hatfalvi, Amber Horvath, Deby Katz, Roshni Kaushik, Mary Beth Kery, Sara Kingsley, Pallavi Koppol, Yasmine Kotturi, Gierad Laput, Tianshi Li, Toby Li, Jirachaya "Fern" Limprayoon, Fannie Liu, Ceci Morales, Tomo Nagashima, Tricia Ngoon, Oscar Romero, Kate Shih, Franky Spektor, Jaemarie Solyst, Françeska Xhakaj, Allan Wang, Sanne van Waveren, Kristin Williams, Naomie Williams, Judith Uchidiuno, Zheng Yao, Qian Yang, Nur Yildirim, Michelle Zhao, and Siyan Zhao. Thank you to my friends from my childhood, college, and grad school years, who have taken interest in understanding how PhDs work, provided boundless support, and lent refreshing and reinvigorating perspectives to my research. I am grateful to the CMU IRB—especially Gail Kusbit—and to the SCS, RI, and HCII staff members who booked rooms, processed reimbursements, and coordinated events, including Peggy Martin, Marian D'Amico, Catherine Copetas, Karen Harlan, Carolyn Buzzelli-Stumpf, Ryan Ries, Patrick Brinkman, Diana Rotondo, and John Davis.

I am fortunate to have embedded myself (at various points in the past few years) in the Mendelssohn Choir of Pittsburgh, the SCS Grad Musical, and Chabad of CMU & JGrads. These communities have made my Pittsburgh life fun, given me lasting friendships, and nurtured both the research and non-research parts of my identity.

I want to thank Xiang Zhi Tan for enriching my work and enlivening my life every day since we met. In my time as a graduate student, Zhi has been my rock through every rejected paper, intense workweek, and personal crisis, my encyclopedia of miscellaneous (and very useful) numerically-oriented facts, my partner-in-crime, and my best friend. There is no one in the world with whom I would rather be celebrating this milestone.

Finally, I want to thank my family. My parents, Shari and Neil Reig, have always encouraged me to explore my interests, given me much-needed perspective in the face of the stress and challenges of research, and reminded me not to take anything too seriously. My little brother, Adam Reig (himself a budding researcher!), is my inspiration in all I do, including this thesis. Thank you to my grandmother, Barbara Kressner, and to the rest of family (including Molly), for unending love and support. Desmond, my cat, came into my life halfway through my PhD, and filled it with cuteness and chaos. Once, as I sat at my desk squinting at Overleaf, he stepped on the D, E, and S keys, in that order—"DES"(mond)—thus attempting to claim this document as his work. Believe what you wish about that.

Contents

List of Figures

List of Tables

List of Abbreviations

HAI	**H**uman-**A**gent **I**nteraction
HRI	**H**uman-**R**obot **I**nteraction
HCI	**H**uman-**C**omputer **I**nteraction
STS	**S**cience and **T**echnology **S**tudies or **S**cience, **T**echnology, and **S**ociety Studies
AI	**A**rtificial **I**ntelligence

Part I

Research Framework

Chapter 1

Introduction

Conversational agents now read and send messages, tell jokes, track package deliveries, and control lights and thermostats. Agent technologies leverage increasingly more hardware to perform increasingly more functions and interface with people increasingly more seamlessly. This rapid evolution brings with it a few conceptual shifts in how we think about socially interactive technologies. First is the changing understanding of the meaning and role of "artificially intelligent agents". In the past ten years, conversational agent design has evolved from reliance on simple voice recognition tools that intake carefully-structured requests and emit mechanical responses to the deployment of widely accessible large language models that mimic human conversation patterns. As we develop these technologies, we see old ideas coming to fruition in new ways and nudge our reality closer to the limits of our collective imaginations. Incremental progression toward the extremes of what might be possible changes what appears possible, thereby shifting what researchers and developers need to consider as future states of the world.

Another shift is that from singular intelligent technologies to integrated ones and "smart environments". Intelligent systems are becoming less and less constrained to single embodiments: voice-activated agents that are typically embodied in smart speakers, for example, can interact with users through multiple physical platforms (such as robots) and control multiple devices in a shared space. Little research has been done to study human-agent interaction (HAI) with intelligent systems that are complex and integrated, not only in terms of their back-end functionality and technical cross-compatibility, but also in how they are designed to communicate with people. An understanding of these interactions requires a careful consideration of what *an agent* (conversational, embodied, social) really is within its immediate social and technological context.

My research centers on *agent identities*, which include the social roles of agent technologies as well as user experience in physically complex, integrated systems. Within the human-computer interaction (HCI), human-agent interaction (HAI), and human-robot interaction (HRI) research communities, as well as in computer science and sociology more broadly, there are numerous definitions and concepts of *what an agent is*. For the purposes of this dissertation, I apply the following definitions:

An **agent** is a computational, interactive, semi-social player with a weak sense of *perceived* identity. An **identity** is the set of characteristics that a designer or a user projects onto an agent and that make it uniquely identifiable to that user.

These concepts are discussed in detail in Chapter 2.

Agent identities mediate interactions between people and embodiments—physical hardware, such as robots and smart home devices—by providing an interface to the hardware and shaping the user's impression of it. They can also mediate between people and services (e.g., by delivering a service to a customer), and between services and embodiments (e.g., by interacting via a robot that is owned by a company that delivers a service, and therefore is part of the service). However, they can also mediate relationships among multiple entities in these categories. For example, if an individual interacts with a single agent across two services (e.g., a grocery store and a medical center), that individual might perceive both services to be more or less trustworthy, intelligent, or social. In this situation, the agent identity mediates relationships among a *single* individual and *multiple* services. If that agent also interfaces with the individual through multiple physical embodiments—such as a kiosk at the grocery store and an interactive autonomous cart at the medical center—then it is mediating the individual's relationship with multiple services *and* multiple embodiments associated with those services.

1.1 Flexible agent embodiment

I use *flexible agent embodiment* to refer to a single agent "identity" or "mind" with a dedicated social presence that can exist independently of a "body" and move between "bodies". In a design exploration of this concept, my colleagues and I explored four "mind-body configurations" for agent social presence (see 1.1)[1]. Agents and robots can be defined to have one "mind" per "body" (*one-for-one)*, following a Cartesian dualism-inspired view of the human model of identity and social interaction. Alternatively, multiple "bodies" could be controlled by one "mind" (*one-for-all*). Agents can also *re-embody*, moving their social presence from one tangible "smart" object (e.g., robot, kiosk, car) to another tangible "smart" object. Finally, they can *co-embody*: multiple agent identities might perform their social presence through the same "body" at the same time. As there is reasonable cost to instantiating multiple, movable social agent presences, future services could deploy a wide array of personalized, branded, and unique agents.

Beyond service interactions, flexible agent embodiment has great potential to enhance human-robot collaborations. Consider, for example, an astronaut who works with an AI in training for a mission. Over the course of several months, the AI learns about her professional and personal history, her areas of expertise and her strengths and weaknesses within them, her health-related needs, her family, her personality, how she behaves under pressure, and her communication style. When she is ready to begin her extraterrestrial mission, that AI embodies the robot who will collaborate with her on the task. Because she is working with a familiar agent (albeit in a different robotic "body"), the astronaut is able to make use of what the robot has learned about her from prior interactions. She can be confident about what the robot knows and does not know, including its understanding of *her own* knowledge and expertise; feel comfortable with the style of communication necessary to exchange information with the robot; and trust the robot to correctly interpret her words and actions. Here, re-embodiment helps

[1]This work was led by another PhD student and was a collaboration among three PhD students and three faculty members. Along with the other two PhD students, I conducted a literature review, ideated on concepts, brainstormed relevant contexts in which to probe agent social presence, developed scenarios, ran all of the studies, analyzed the data, and wrote up the findings.

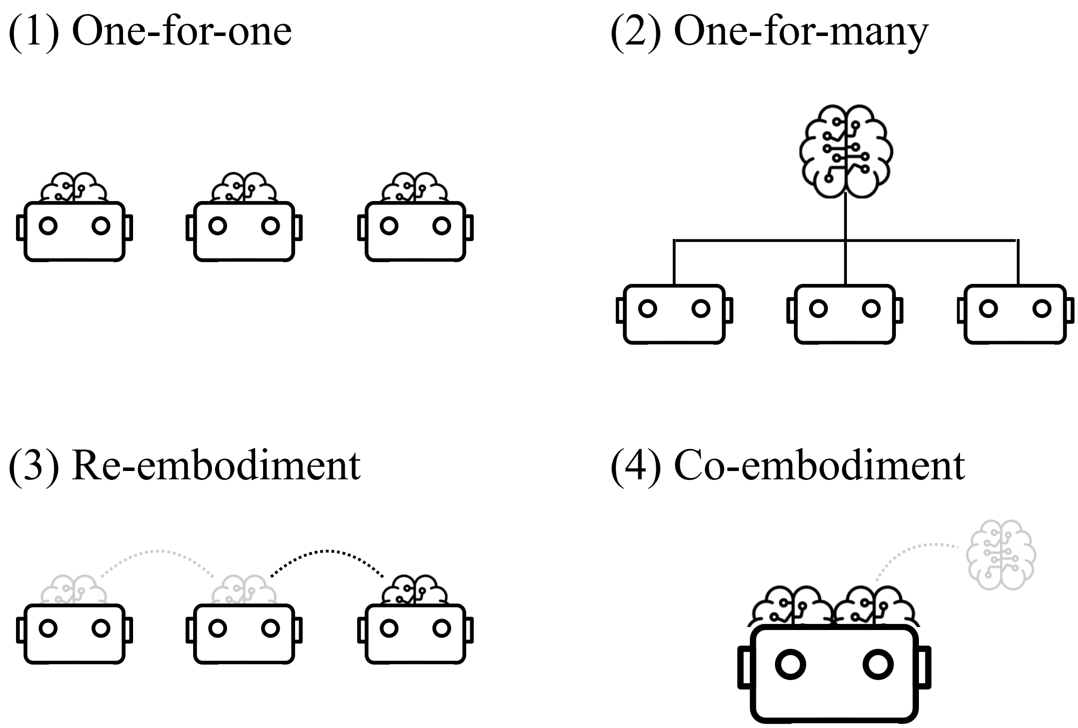

FIGURE 1.1: Diagram of four paradigms for embodying agent identities in robots ("social presence options") from Luria et al. (2019).

a human-robot team be resourceful in establishing and taking advantage of common ground, encourages a trusting relationship, and utilizes social presence and customization to facilitate smooth interaction.

1.2 Agent identities as mediating entities

Within the fields of HRI and HAI, *mediation* may be assumed to refer to *conflict resolution*. In HCI and science and technology studies (STS), *mediation* takes on a broader, more philosophical meaning. Ihde's postphenomenological approach to technology evolved the idea of technolog*y* as a general, abstract category into the concept of technolog*ies* as distinct, literal entities with, through, and around which people interact (Ihde, 1990). Verbeek (2006) built upon this progression with the theory of technology mediation: the principle that technology is not just a product or a result of human creation, but something that shapes real-world interactions.

Verbeek (2015) describes three dimensions along which technology can mediate. The first, *types of relations*, addresses the directionality of the human-technology-world relationship in a direct application of Ihde's work (Ihde, 1990). Relationships can fall into the category of embodiment, in which technology enables a novel interaction within the world (e.g., people communicating by telephone); hermeneutic, in which technology allows humans to read their world (e.g., an MRI); alterity, in which the human and the technology are the focal point and the world provides the context (e.g., a person talking to a robot); or background, in which the technology is part of the context along with the world (e.g., a smart environment with light and temperature sensors). The second dimension, *points of contact*, regards the nature of the impact of the technology on people—they interact physically, cognitively, or emotionally. The third is *types of influence*, which is the amount of control that the technology has over the human's experience, i.e., whether it forces a particular action or interaction (e.g., a subway turnstile), or simply suggests its intention without forcing it (e.g., a fuel efficiency meter on a car). Verbeek's work suggests that designers can—and in fact, are ethically obliged to—either *anticipate* the ways in which what they are creating will serve to mediate people's interactions with each other and the world, or expressly *design* with a specific mediation intent in mind.

Robots as mediating technologies

The closest theoretical explanation of technology mediation as applied to *agents* as they are considered in this thesis comes from Kubo (2010). Kubo, drawing on the same central concept underlying Verbeek's thesis that technologies actively shape interactions among other entities in the world, describes technology mediation as "boundary-crossing" in various forms. He identifies three aspects of technology mediation: *connecting heterogeneous entities*, *connecting the material and the conceptual*, and *facilitating interaction between different perspectives*. Kubo explains the latter two aspects via the example from Latour, 1999 of a speed bump on a university campus as a mediating technology[2]. A speed bump takes the goal of *slowing down cars to protect pedestrians* and physicalizes it into something concretely impactful for the driver: drivers will slow down if they want to spare their cars from damage, and someone crossing the road on foot will be safer as a result. In this way, a speed bump externalizes a conceptual goal and makes it a material one. It also mediates among perspectives: rather than drivers and pedestrians directly fighting over their conflicting needs, they compromise over, and are constrained by, the speed bump. Kubo does not explain connecting heterogeneous entities through this example, but the reader can extrapolate: interactions between cars and pedestrians—different categories of moving things—are mediated by the speed bump.

Kubo performed a cultural anthropological analysis of the engineering process and adoption of the SONY AIBO robot dog (*Aibo* n.d.), which was developed throughout the 1990s, productized in 1999, and sold to tens of thousands of households in the early 2000s (Taub, 2006)[3]. AIBO incorporated new technical principles from Brooks' subsumption architecture and behavioral robotics (Brooks, 1991), but creators also had to consider how its consumers' mental models of pets should shape its design to make it marketable. This gave it promise as a kind of technology with a novel cultural role that had never been realized before. To make the most of its technical capabilities and thoughtfully define its position as a companion technology, its development team included people with diverse engineering and scientific backgrounds. Kubo argued that AIBO connected heterogeneous entities by building networks among practitioners from different professional disciplines. Moreover, it connected the material and the conceptual by bringing intellectual contributions to life in concrete terms in the form of a social

[2] Latour also used the speed bump to describe mediation as the translation of disparate conceptual goals into a common material; Kubo's contribution is in elaborating on it, and in describing it in terms of his three aspects of mediation.

[3] The last release of the original edition of AIBO in Japan was in 2006. In 2018, SONY AIBO resumed production and sales of the robot dog, including in the U.S. (Sorrentino, 2018).

Technology	Connecting entities	Bridging conceptual and material	Connecting perspectives
Speed bump (as in Kubo, 2010, c.f. Latour, 1999)	Hedged between physical entities of cars and pedestrians	Introduces and concretizes a common goal: drivers protect their cars, pedestrians move more safely	Balances needs of drivers (to drive cars), of pedestrians (to avoid cars), of university (to maintain safe roads)
AIBO (as in Kubo, 2010)	Joint accomplishment of engineers, scientists, customer service, UX practitioners	Actualizes intellectual technical advancements as a reality	Balances conflicting opinions about design (e.g., efficiency vs. psychological effects)
Migrating agents	Hedged between physical entities of people, hardware, and collectives (i.e., companies, services)	Actualize visions of automation; actuates embodiments; facilitate ascription of identities to embodiments; deliver services	Balance conflicting stakeholder perspectives (e.g., by controlling privacy and the flow of information)

TABLE 1.1: Analogical application of the three aspects of technology mediation from Kubo (2010) to the examples of a speed bump, the AIBO robot, and migrating agents.

machine. It also facilitated interaction between different perspectives because it served as a focal point for stakeholders' debates about design and functionality[4]: it mediated among different, sometimes-conflicting viewpoints about AIBO's role. These perspectives centered the AIBO as a particular, but also abstracted beyond it to the concept of technologies that occupied (or could occupy) the role in which AIBO served. In terms of its reception, AIBO mediated its users' ("owners"') interpretations of its behaviors and role in that it embodied the formulation of these interpretations based on multiple factors: users' sensemaking about its engineering, its actual engineering, their interactions with other "owners", and their own preexisting cultural practices and frames (e.g., mental models of pets).

These theories provide a framework that I apply for my interpretation of mediation as it applies to agents and their relationships to individual people, hardware embodiments, and services (see Table 1.1 for an application of the Latour (1999) notion of boundary-crossing and the Kubo (2010) interpretation of it to migrating agents). Verbeek said that "ultimately, it is not *things* that are to be designed, but rather the *interactions* between humans and things" (Verbeek, 2015, p. 26); one could envision *all* of HCI as a study of mediation, with technologies as anchor points along paths connecting humans to each other and to their environments. Agents, and particularly migrating agents (see Sections 1.1 and 2.1.3), take up a unique position in this network when we view them primarily as user-constructed (or other stakeholder-constructed) identities

[4]Kubo provided the example of the tension between the inclination to view the robot as analogous to a living dog in that it was always either awake or could be woken up to interact, and the practical constraint of its battery life. AIBO's charger was designed to mediate this tension: it was originally a box that the robot would be put in to recharge, but because locking the dog away felt too much like "killing" it, the charger was eventually designed as a platform that would not block the owner's view of their dog but instead make it appear present and "sleeping".

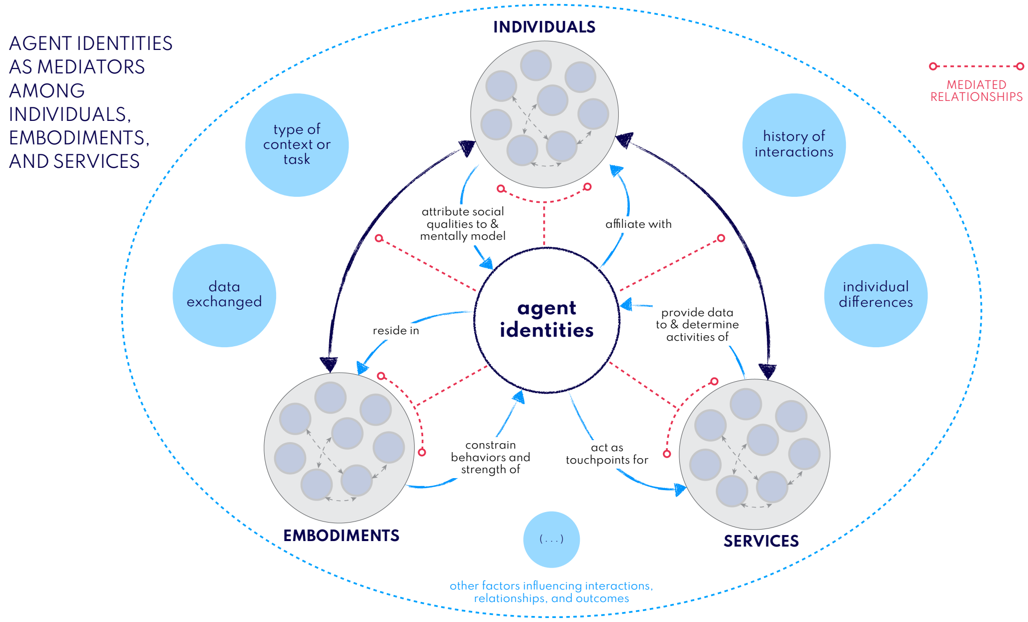

FIGURE 1.2: Conceptual model of agent identity as a mediator in relationships among individuals, embodiments, and services.

and are agnostic about their form. Because they are not constrained by a particular embodiment or to a particular context, their role as mediators can go beyond mediating between entity and entity, vision and reality, and perspective and perspective: they can also connect *other* technological entities that are *already* doing this kind of mediation (e.g., robotic products, technological services).

1.3 Research overview

This dissertation presents a three-part exploration of aspects of the agent identity model. Part I provides an overview of the research framework for this work. In Part II, I describe two studies in which agent identities are positioned as *touchpoints in service design*. Using research-through-design approaches, I explored what it might mean for a single agent identity to manage an individual's relationships with various services (e.g., stores, healthcare) by way of embodying and "re-embodying" robots that are operated and maintained by the service.

Part III discusses two studies that shed light on agent identity as a mediator of interactions across *multiple physical embodiments and multiple domains* (which are, arguably, "services" in the context of human-robot collaboration). It first describes an empirical investigation of how agent identity persistence impacts perceptions of a multi-embodiment system following a breakdown. It then reports on an empirical study that examines the impacts of agent association, expertise, and system narrative perspective on factors related to trust, performance, and social perceptions of AI teammates in a task-based collaborative setting.

Part IV focuses on *social roles of agent identities* among multiple individuals in everyday life. It describes a Wizard-of-Oz study in which an agent relied on the notion of the implied presence of an absent third party to give a participant good-quality (acting in the participant's interest) or poor-quality (acting in the absent third party's interest) information. This manipulation had a greater effect on people's perceptions than whether the agent was embodied virtually, physically in a robot, or not at all. It then describes a story completion study in which participants wrote creatively about future smart home interactions. This work revealed gaps between smart home agent and device capabilities and actual use as well as differences between how people actually interact with the devices in their spaces and the interactions that they envision when asked to imagine future smart home scenarios. Finally, Part V contains reflections on the scientific knowledge, methodological innovations, and design implications produced by the work in this thesis.

Chapter 2

Background and Related Work

Portions of this chapter were previously published in the scientific articles:

Samantha Reig, et al. (March 2020.) "Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot". In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*.

Samantha Reig, et al. (February 2021.) "Perceptions of Agent Loyalty with Ancillary Users". In *International Journal of Social Robotics*.

Samantha Reig, et al. (March 2020.) "Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot". In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*.

Permission to reproduce the text as part of this dissertation is included in the original publication's copyright.

This literature review draws on five areas of research that provide a theoretical framing for my thesis work. First, I present relevant theoretical and empirical work on human-agent interaction. A historical perspective on agents as software identities formulates a working definition of the term "agents" as I apply it in my research (as computational, interactive, semi-social players with a weak sense of perceived identity) and lays the groundwork for the concept of flexible agent embodiment in HCI research and service design. Second, I discuss several studies of how agents and robots can affect group dynamics (i.e., how they function in complex social environments) and serve as interfaces to multi-interface systems (i.e., how they function in complex physical environments).. These provide insight to the study of multi-person, multi-agent, multi-interface interaction. Third, I review literature on robot embodiment. Fourth, I synthesize work from literature on symbolism, self-extension, and asynchronous interaction. Together, these concepts contribute context for research on how agents may mediate interactions between individuals over time. Finally, I provide a brief overview of robots in service design, which motivates much of the methodology and scopes the contributions of the studies in this thesis.

2.1 Agents as software identities

2.1.1 Agents

The term "agent" has historically been used in computer science to refer to various kinds of entities, from something as simple and machinelike as a software process spanning a few lines of code to something as complex and interactive as the Pepper robot (Pandey and Gelin, 2018). Prior work notes the historical lack of a single definition (Jennings, Sycara, and Wooldridge, 1998; Franklin and Graesser, 1997), so it is useful to begin by clarifying what it will come to mean for the purposes of this thesis and why this term is appropriate.

HCI research uses both technical and social conceptualizations of agents. On the technical side, definitions usually reflect *automated software processes* that are "situated in some environment" and "capable of flexible autonomous action in order to meet [their] design objectives" (Jennings, Sycara, and Wooldridge, 1998), and "machinery for sensing a user's activity and taking automated actions" (Horvitz, 1999). The literature on multi-agent systems distinguishes agents from one another in terms of their purview: they are specialized computational actors within a larger architecture in which "there is no system global control; data are decentralized" (Sycara, 1998). In this *technical* perspective, what exactly within a complex system is considered "an agent" by the developer is not always the same as what appears to be an agent to users on the front end. On the *social* side, definitions reflect a more human-focused viewpoint and motivate more user-centered (and arguably more liberal) usage of the term. Lieberman and Selker (2003) provided the criterion that "the machine can be considered an agent if the best way to explain its behavior is by analogy to the agential role that humans play". Maes said that "the metaphor used is that of a personal assistant who is collaborating with the user in the same work environment" (Maes, 1994).

These two ways of thinking about *agents* are also reflected in the distinction made by Wooldridge and Jennings (1995) between something with *weak agency*, which is capable of autonomy, perception, and proactiveness; and something with *strong agency*, to which humanlike traits, such as mental states and emotions, can be attributed. However, the two categories are not quite at odds; real systems draw on both. The formalization of *agents* in the Open Agent Architecture (OAA) (Martin, Cheyer, and Moran, 1999) allowed for any software process to be labeled as an agent (or not) based on its relationship to an existing set of processes that were already called agents. Much of the work on the OAA eventually led to the development of the CALO (SRI, n.d.) project and eventually Apple's Siri. Though the Siri system, when traced back to its roots, could be said to comprise *multiple* AI agents, the total system takes on a singular identity that is considered "an agent" by users. Recent work notes that what constitutes an agent remains a case-by-case design choice (e.g., Chung et al., 2019).

This thesis will usually refer to "agents" where others might refer to "conversational UIs", "AI agents", or occasionally even "systems". When talking about the agents of the present, I will sometimes use a word that most accurately describes what kind of agent a technology is; for example, Siri is an agent that is a voice interface or a virtual assistant and Pepper is an agent that is a robot. Because questions about the meaning of embodiment, physical design, and social behaviors of agents are at the very core of this work, I will usually characterize future technologies with different, changeable, or unknown embodiments as "agents" without differentiating them by capabilities or form, and I will sometimes elaborate with language to reflect the context (e.g., "home robot" or "personal agent"). At a high level in this work, ***agents* are systems that complete (some or all) tasks autonomously, respond to the activity and/or input of other agents, can communicate legibly to humans, and are viewed by people as an entity to which at least minimal social qualities can be attributed.**

2.1.2 Identities

The Wooldridge and Jennings (1995) application of weak vs. strong identity to AI agents provides a theoretical jumping-off point for understanding what constitutes identity in agents. More recent work investigates how the construct of a non-human system's *identity* might arise from specific interaction cues. One approach leverages Levels of Abstraction (Floridi and Sanders, 2004) to explain how people perceive robot agency. Levels of Abstraction (LoA) have to do with the perspective from (or level at) which a particular concept is studied or specified. Before attempting to understand a construct such as moral agency, social agency, or identity, one must specify the set of parameters that would constitute the presence, absence, or degree of the construct for the LoA of focus.

For robots, LoA are related to familiarity with the technology and mental models of the robot. Jackson et al. (2021) explain that robot agency and identity are things that must be examined and measured in terms of the observables at a particular level of abstraction. For instance, a developer may not believe that a robot they helped to create has any agency or an identity because they are familiar enough with how it functions, what and how it learns, and why it takes the actions that it does to be able to fully analyze its behaviors and faults in terms of specific logical flows. To a person interacting with the robot, however, the robot might be an agent because it appears to make decisions on the fly, presents as having a set of personality characteristics that cannot be described entirely in terms of enumerable behaviors, and violates some expectations (see Forlizzi, 2007; Forlizzi, 2008 for detailed accounts of how robot vacuums in homes can become *social products*). Names, speech (including the use of pronouns and possessives), movements, behaviors, and physical appearance are all *robot identity observables* at the *user's* LoA (Jackson et al., 2021). Determining what exactly constitutes agent and robot identity from a human psychology perspective and understanding the role and impact of agent identities in shaping interactions among other entities are distinct, but complementary, research problems. My use of "agent" (as described in the previous section) focuses on the *user's* level of abstraction, and assumes the existence of some notion of "identity" at that level.

2.1.3 Agent migration: Theoretical and psychological roots

The concept of software intelligence migrating across physical platforms was originally proposed by Duffy et al. (2003). Their Agent Chameleon framework proposed an architecture in which software agents could move between virtual and physical environments as well as mutate (e.g., by gaining or losing a physical feature) within an individual environment. It positioned an Agent Chameleon as an autonomous and portable entity with a set of capabilities (some of which are predefined and unchanging, others of which adjust according to the environment) and platform-dependent social abilities. In order to migrate, an agent must contact a hub that creates a copy of the agent, including its mental state, in the desired arrival location. The old agent is then erased. The framework also emphasized the importance of equipping agents with basic survival instincts: if an agent perceived that an external force (such as a dying battery) would soon cause it to be unable to function, it could either migrate or "save" its internal state to storage. Of course, the scope of the Agent Chameleon Project was mostly hypothetical, as the technology to create migrating software intelligence capable of re-embodying multiple physical and virtual environments and having smooth interactions with humans did not exist in 2003 the way it does today.

Martin et al. (2005) framed *agent migration* as a matter of how the issue of mind and body has manifested in the artificial intelligence community. These authors posited that that an agent's identity relies on how it is perceived by its users and can be broken down into individual cues. To validate this idea, the authors performed an experiment in which 31 participants were shown migration across virtual avatars, with continuity of the agent cued as either a shared feature (such as a hat or glasses), a shared color scheme, a shared set of markings, or a shared form. The cue with the highest similarity rating was features, followed by form class, markings, and color. The first experimental implementation of this type of migrating software intelligence was the proof-of-concept ITACO system by Ogawa and Ono (2008). In one of the ITACO studies, participants who interacted with an agent that spoke to them, migrated to a robot, and spoke again had an easier time understanding the robot and were more likely to respond to the robot's implicit request than participants in the control condition (in which the agent did not migrate to the robot before the robot spoke). In a second study, participants observed an agent migrating from a wearable computer screen to a lamp. They were then asked to turn off the lamp, and they reported a sense of loss in doing so. This project was motivated by the idea that interactions between humans are emotional as well as functional, and that conversation is pragmatic rather than literal. Therefore, having the same agent appear in various objects in an environment accommodates the emotional nature of human interaction and thus makes the interactions between the humans and the artifacts more natural. My thesis work is similarly motivated by this emotional aspect of human relationships and the pseudo-emotional potential of agent identities in human relationships.

The LIREC (LIving with Robots and intEractive Characters) project (Paiva, 2017) positioned a migrating intelligence as a companion technology that could provide continual social support while offsetting the power costs of carrying a physically embodied agent from place to place (Kriegel et al., 2011). Kim et al. (2013) conducted an experiment to test people's psychological boundaries of robots. In a controlled study, robots asked participants to perform a physical task with the robot by moving it to the side. The robot said, "Slide me slightly," and the researchers' goal was to test where and what the participants perceived the "me" to be. In four different conditions, the experimenters told the participants that the robot was composed of various combinations of cart, table, and robot, differently equipped each time. After the interaction, participants drew "me", and experimenters counted how many participants drew the robot, how many drew the table, and how many drew the robot together with the table. The researchers concluded that recognition of "me" is controlled by the level of uncertainty or ambiguity in the instructions (i.e., the framing of what is and what is not called a robot). Several of the studies in my thesis work rely on this notion of framing in that they utilize language and design cues to communicate what is "the agent" in probing agent identity.

The study of agent migration has been extended to psychology research settings through determining children's concepts of migrating intelligences (Syrdal et al., 2009) and probing potential configurations of embodiments and intelligences (Koay et al., 2009). Research in situated laboratory contexts has exhibited prototypes of migrating intelligence in robots inhabiting mock smart homes (Koay et al., 2016) and compared the effects of identity (i.e., behavior and personality) migration and data migration (Tejwani et al., 2020). Both Koay et al. (2016) and Tejwani et al., 2020 suggested that a persistent "identity" of an AI over time and embodiment is a crucial benefit of migrating intelligence.

2.2 Agents as social players and intermediaries in groups and teams

Prior work on agents and robots in group interactions informs the aspects of the conceptual model of agent identity (Section 1.2) that depict (1) the relationship between agent identities and individuals and (2) agents as mediators among individuals.

HRI researchers have begun to position robots that work with human teams as teammates rather than as tools. Ma et al. (2018) articulated a number of considerations for developing a general theory of human-robot teaming. Among the points they emphasized was the importance of determining how to assign task work; examples include function allocation (dividing and then allocating tasks based on the abilities of the agents involved) and interdependencies (see Johnson et al., 2014). The Shared Mental Models computational framework for human-robot teaming has robots share knowledge about task progress, teammates' statuses, and changes to the environment (Scheutz, DeLoach, and Adams, 2017). This model was evaluated in two studies (Gervits, Fong, and Scheutz, 2018; Gervits et al., 2020), both of which showed that sharing mental models improved task performance but did not affect human teammates' subjective perceptions of workload and situation awareness. This prior research provides a foundation for defining the roles of multiple agents that collaborate with humans in task-based settings and suggests that agents that share information can improve humans' performance. My work draws from and aims to contribute to this body of work.

Several works have studied how the social behavior of robots in a multi-robot setting affects people's perceptions of the robots. In one study Tsujimoto, Munekat, and Ono, 2013, participants perceived robots that interacted with them more favorably and were more likely to take their recommendations relative to robots that did not interact with them. Relatedly, Sembroski, Fraune, and Šabanović (2017) found that people complied with a robot's request that conflicted with an experimenter's when the robot was an in-group member and the experimenter's authority was low. Fraune, Šabanović, and Smith (2017) simulated a competitive task in which teams were comprised of two humans and two robots each, and found that in-group members were seen more positively than out-group members, whether they were humans or robots.

Fraune and colleagues made several discoveries relevant to group HRI/HAI in smart environments in their work on robot *entitativity*, or how much a group of robots are perceived as a single entity as opposed to multiple entities (Fraune et al., 2020; Fraune et al., 2017). In one study, participants perceived entitative robots to be more socially threatening than diverse robots or single robots, and robots in a diverse group were perceived as most intelligent (Fraune et al., 2017). In another, perceiving robots as entitative was associated with more willingness to interact with them and more positive perceptions overall, but robot-robot interactions did not increase perceptions of entitativity (Fraune et al., 2020). Beyond this, several group effects originating in human social psychology have been found to translate in some capacity to HRI settings, including conformity to the popular opinion (Salomons et al., 2018), effects of passive presence on honesty (Hoffman et al., 2015), and ripple effects of prosocial behavior (Correia et al., 2019; Strohkorb Sebo et al., 2018).

In some situations, introducing multiple disembodied or fluidly embodied agents can overcomplicate the interaction. Chaves and Gerosa (2018) had participants interact with either one omnipotent chatbot or multiple "expert" chatbots (with expertise in nature, culture, and shopping) in a travel-planning scenario. In this task-based situation involving a single user, the only notable difference between the conditions was that participants reported being more confused when interacting with multiple chatbots. The research described in Chapter 3 also explores the notion of designing multiple agents and tells somewhat of a different story: in a service setting, assigning one agent to each user can give people a sense of personalization and comfort (Reig et al., 2020). The work that I describe in Chapter 6 addresses this concept in a task collaboration scenario.

2.3 Embodiment

Underlying the work on re-embodiment, which is described in Section 1.1 and frames my thesis, is a long history of research on how agent embodiment impacts interactions and perceptions. In general, the design and appearance of any agent influences the way people perceive it. For example, people may rate a robot as more knowledgeable about dating when they perceive it to have a gender that matches their own (Powers et al., 2005). Most findings regarding the differences in interaction outcomes among robots, avatars, and voice agents for non-physical tasks point to a positive effect of having a physical embodiment. In one study, performance on a memory task was better after an interaction with a virtual robot than with a co-located, embodied robot, but ratings of sociability, responsiveness, competence, trustworthiness, and respectfulness were higher for the embodied robot (Powers et al., 2007). In a decision-making scenario, people felt more attachment (a combination of liking, preference, and negative reaction to potential loss) to an embodied social robot than a virtual one (Wang and Rau, 2018). Physical embodiment is also associated with an increased tendency to anthropomorphize. Participants in one study found a co-located robot more engaging than a co-located virtual agent, a remote virtual agent, and a remote robot, and they interacted in more anthropomorphic ways with the robot than with the agent (though they anthropomorphized both) (Kiesler et al., 2008). Additionally, people may more freely make conversational or "in-the-moment" anthropomorphic assumptions about specific robots than about robots in general (Fussell et al., 2008). Some work suggests that a physical robot body can also affect people's behavior by way of its mere presence: in one study, having a robot monitor in the room led people to curb their cheating behavior just as much as having a human monitor (Hoffman et al., 2015).

Work on embodied conversational agents (ECAs) (e.g., Cassell, 2009; Cassell et al., 1999; Cassell et al., 2000) suggests that the value of embodiment for developing rapport centers on physical and behavioral design features: having an anthropomorphic form, gesturing, and nonverbal backchannelling. Other research has found key differences between human-human dialogue and human dialogue with embodied conversational agents (ECA) and conversational user interfaces (CUI): people do not speak to conversational assistants in the same way that they speak to humans, but they do make certain social attributions to agents similarly to how they make social attributions to humans (Luger and Sellen, 2016). A review of experiments on the effects of physical robot embodiment and presence in HRI (Li, 2015) concluded that, overall, physical presence or lack thereof has a greater impact than robotic platform and that robots that are embodied and present are more persuasive and viewed more positively than virtual avatars or robots that are embodied but not present. Another survey of several dozen papers on robot embodiment found that its impact is mostly positive on both task performance and perceptions of the agent (Deng, Mutlu, and Mataric, 2019).

Designers have considered how task and context might guide decisions about how to embody a robot. One approach (Deng, Mutlu, and Mataric, 2019) characterizes designing embodiment as following a design metaphor to some level of abstraction. When deciding how to create an embodiment for a robot, designers need to determine what the metaphor is (what in the world it is meant to emulate) and how abstract or literal its implementation should be (how closely it needs to adhere to the metaphor in order for its affordances to be understood). For example, if it is extremely important for people to understand that a particular robot can speak, that robot may benefit from having a mouth that closely resembles that of a human. In contrast, if accurate perceptions of speech affordances are not as critical and the robot is meant to have a sleek and elegant look, then a mouth that is more animal- or machine-like or altogether absent may be more appropriate.

Most empirical studies of agent embodiment employed tasks conducive to face-to-face conversations in which the user was not distracted by other factors in the environment. Chapter 7 describes a study in which I investigated how findings from this body of work translate to scenarios that place external demands on the user's attention, and where the agent identity sends stronger signals than the embodiment.

2.4 Implicit mediation over time

One of the ways in which agents can mediate interactions between people is by transmitting messages from one person to another. This is often explicit: "Siri, send a text to my daughter to tell her to please walk the dog." It can also be implicit: imagine one member of a household arriving at home, saying, "Alexa, play" (without knowledge of what music had recently been paused), being met with a number from a musical soundtrack, and becoming newly aware that whichever member of their family last asked Alexa to play music is fond of that musical. Robots—embodied agent identities—are particularly suited to implicit mediation because they can draw on more cues. For example, a robot's social eye gaze has been found to mediate interactions among members of a group (Admoni and Scassellati, 2017), group dynamics constructs like participation (Gillet et al., 2021), and the establishment of roles (Mutlu et al., 2012). Additionally, body orientation has been shown to influence physical group dynamics (i.e., configurations and orientations, see Vázquez et al. (2017)).

Most research on how robots mediate relationships and interactions among people is concerned with synchronous interaction, or focused on the here-and-now. There is, to my knowledge, little work on *asynchronous* interaction *between* people *via* agent identities and robots. However, prior work from psychology and HCI inspires and provides context for an exploration of robots as mediators of relationships between individuals over time, which is a theme throughout this thesis.

2.4.1 Self extension and symbolism

According to Belk (1988), self extension is a phenomenon in which individuals project aspects of themselves beyond the body and into inanimate objects. Belk claimed that "We learn, define, and remind ourselves of who we are by our possessions... Our accumulation of possessions provides a sense of past and tells us who we are, where we have come from, and perhaps where we are going" (Belk, 1988). Research has investigated self extension into inanimate objects: in Kiesler and Kiesler (2005), participants who designed a pet rock for themselves saw the rock's "personality" as more similar to their own and were less willing to productize the rock than participants who designed a pet rock to sell to someone else. Work on self-extension into robots has focused on physical interaction: Groom (2010) describes the effects of robot autonomy, mediation, form, artificiality, and the operator's prior experience with the robot on self extension into teleoperated robots (Groom, 2010). There is little work on how the self extends to robots that themselves project somewhat of a "self" through the connotation of a weak identity (i.e., social robots).

Belk also argued that the extended self is not made up of only concrete objects. It can also include money, pets, other people, and body parts, (Belk, 1988), and can take on a different shape in digital spaces based on objects and experiences from those spaces (Belk, 2013, see also Cushing, 2011). Gosling et al. (2002) also orbits the concept of how the self is projected beyond the body, but considers the perspective of the receiver. Gosling found that in rooms, people leave *identity claims* and *behavioral residue*. Identity claims are "symbolic statements made by occupants" (Gosling et al., 2002, p. 3), such as posters, souvenirs, or the color of paint and furniture. They communicate what a person *is*. These can be self-directed (serve to reinforce the occupant's own identity without concern for the judgments others will make) or other-directed (serve to craft an image of oneself for others to use in making judgments about them). Behavioral residue is "the physical traces of activities conducted in the environment" (Gosling et al., 2002, p. 4). They communicate what a person *does*. This residue can reflect activities that have happened or will happen in the immediate space (interior residue, e.g., an open bottle of wine and a board game; an organized CD collection), or activities that are conducted outside the environment ("exterior residue", e.g., a pair of skis; a subway card). In two field studies—one in an office, and one in personal living spaces—Gosling and colleagues examined how and with what degree of consistency people make judgments about others based on their personal spaces, and found that people can indeed reasonably accurately infer others' personality based on what they leave behind.

2.4.2 Situated delayed communication

Several works have explored how situated interfaces might support asynchronous interaction among family members. Sellen et al. (2006) prototyped and studied the *HomeNote*, an asynchronous messaging system. The HomeNote had a function similar to that of a group chat and a form similar to that of an iPad, but was intended to be placed in a specific location within the home. It was designed specifically with *human-to-place* communication in mind: members of a household could send situated messages intended for whoever next entered a *place*, rather than for a specific *person*. For example, a parent on their way back from work could broadcast "Can someone pick me up from the train station?" to the family fridge, and then whoever next passed by the fridge and was available could respond, e.g., "Went to pick up Mom—Adam". Advancements in smart home technology have paved the way for research into devices that record and track household activity and brought to light issues about privacy, social norms, and interpersonal boundaries. For example, Singhal et al. (2018) created the *Time-Turner*, a set of three drink coasters that could be used to visualize and interact with video data from an always-on camera that recorded family activity in the home for later viewing. In a study, participants enjoyed witnessing real, forgotten moments from their and their family's pasts after the fact; they found meaning in being reminded of these moments and reliving them accurately (rather than through the distorted lens of memory). However, they also had concerns surrounding privacy and consent among family members (especially as children get older), having a record of moments that they may not want to remember (e.g., embarrassing moments), and embodying sensitive information in a type of object that is often used by guests (coasters).

2.4.3 Asynchronous HRI

Previous research has investigated asynchronous interaction with robots (i.e., between a human and a robot). Marquardt et al. (2009) described a prototype of a Roomba that allowed its users to leave situated messages as commands. Messages left to the robot by the human were of four types: instructions (tasks, exceptions, navigation); context information (environment and location); training; and conditions. The robot could leave the human messages in the form of status, observations, requests, and traces. Messages were sent through RFID tags which could be dropped by the human or the robot and then read by the recipient. Young and Sharlin (2006) proposed the idea of a mixed reality integration environment (MRIE) in which robots could, through augmented reality, leave "thought crumbs" (icons that served as status signals), "bubblegrams" (cartoon speech bubble-like messages to represent the robots' thoughts), and decorations (for fun and expression) in the form of visual augmentations to the scene that the human user viewed.

2.4.4 Sharing bots

Seering et al. (2020) deployed a community-owned chatbot ("BabyBot") into an existing online gaming community on the livestreaming platform Twitch. During several streaming sessions over three weeks, the bot interacted with the community: it inquired about the livestreamer and about the other players, responded to requests for information about its state, and provided commentary about itself and the livestream. The researchers designed the bot with the intention for it to be "raised" by its community, just as a child is raised by its family. They developed an algorithm that allowed it to "grow" over time, changing states from a baby to a toddler to an adolescent to an adult. Over the course of the approximately thirty hours of interaction that the bot had with its users, people engaged in initial sensemaking about the bot by asking it questions. They then tested its limits through humorous aggression[1]. The bot facilitated interactions between humans and itself, but also between other humans, who discussed its behaviors and reacted to its sometimes-odd or not-entirely-parsable comments. Through these interactions, it developed personal relationships with individual community members and became a member of the community.

2.5 Agents and robots in services

The presence of agents and robots in service environments permits a new touchpoint for personalized service. Research shows that people increasingly prefer a single point of contact: customers wish (and expect) to interface with one agent that is knowledgeable about all touchpoints and is situationally and temporally aware (Rapp et al., 2017). This is inherently difficult for human agents, but AI can allow a service to craft personalized experiences that go beyond what people alone can achieve, fostering human-agent service relationships that do not necessarily mimic human relationships. Companies have begun to leverage this, addressing design for the use of multiple voice assistants on the same device (Bohn, 2019; Baldwin, 2019).

Personalization is a key aspect of a user's relationship to a service and has been said to be the most important variable in determining perceived service quality and customer satisfaction (Mittal and Lassar, 1996). HCI research into theories of user personalization of the appearance of computers and phones suggests that while users can apply personalization to their devices of their own accord, features to enable personalization can also be built into the design of the device (Blom and Monk, 2003). Recently, HRI researchers have designed robots with the explicit purpose of personalization and customization of physical appearance and behavior such that the same base platform can be used for numerous projects (Suguitan and Hoffman, 2019). Critically, personalized experiences can also increase loyalty by way of enhancing satisfaction and trust (Ball, Coelho, and Vilares, 2006).

Trust and personalization are often intertwined for robots. Research has shown that a single error can impact humans' trust in the robot, especially in critical situations (Robinette, Howard, and Wagner, 2017). Similarly, a robot's mishandling of personalization may have irreversible effects on a human-robot relationship; for example, a hospital robot that does not provide a patient with their desired level of privacy may destroy trust in that robot, and perhaps in the hospital. Fortunately, personalized interactions with a robot can also be beneficial. In a field study of long-term interactions with a robot embedded in a workplace, incorporating discussion of personalized topics like food preferences, frequency of use, and prior service breakdowns increased rapport and cooperation with the robot as compared with discussing social, but not personalized, topics (Lee et al., 2012).

[1]This phenomenon of "testing" a system to see how far it can be pushed has also been observed in human-robot interactions in public spaces: researchers have interpreted people's "bullying" of robots as being driven by curiosity (Salvini et al., 2010; Brščić et al., 2015).

There is also a demand for personalization: owners changed Roomba appearances to express identity or to fit in in the home environment (Sung, Grinter, and Christensen, 2009), and potential users of elder care robots placed a high value on the affordance of robot personalization to meet patients' particular emotional and physical needs (Moharana et al., 2019). There has been limited work on design guidelines for adaptive robotic services. Lee and Forlizzi (2009) augmented the conventional service blueprint with a *line of adaptivity*, which describes both changes in the service and changes in the user through repeated interactions.

These research efforts, and the majority of work in designing for personalization, have focused on personalization for a single user and had little regard for the surrounding social context. My research considers these issues in the context of agent identities that interact with multiple people (within and across space and time) and take on multiple embodiments.

Part II

Agent Identities as Touchpoints In and Across Service Interactions

Chapter 3

Exploring Personalized Interactions with Fluidly-Embodied Service Robots: User Enactments Study

Much of this chapter was previously published as the scientific article:

Samantha Reig, et al. (March 2020.) "Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot". In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*.

Section 2.5 discussed how customers increasingly desire a single face for the various aspects of their service interactions. However, in many settings, the specialization of human service workers is practical and necessary: they cannot be expected to have the skills, expertise, or interest to perform these multiple disparate tasks as part of their jobs. Here, agent identities could re-embody and co-embody in operating alongside service workers, providing the service-side stakeholders with computational, data, and physical task support while providing customers with a familiar "face" and a sense of consistency across service touchpoints. They could also represent the service to the customer in any number of physical contexts and through multiple interfaces. In this way, they would mediate the *human-service-embodiment* relationship.

We designed a study based on four open-ended research questions intended to inspire and guide scenario design and analysis surrounding a broad, but structured, exploration of this concept. Our first research question pertained to the social norms of human-robot group interactions in service contexts:

RQ1: How should a robot personalize its performance of service with multiple users? How does context influence this?

We also explore the novel question of how multiple social agents should interact through the same physical platform (co-embodiment):

RQ2: How does co-embodiment impact people's perception of the service robot experience?

FIGURE 3.1: Our service robot prototype. The images displayed on the screen changed as different agents embodied the robot at different times.

Re-embodiment gives social robots the opportunity to make use of multiple individual, personalized agents that tailor their behavior to their primary users. This opens up questions about how people develop relationships with agents, the robots they embody, and the services with which they interface:

RQ3: How does a sense of personal connection to a robot's intelligence influence trust in that robot and feelings about the services it helps to provide? What is the social role of a universal personal agent?

Finally, re-embodying agents can interact with people through different robots, in different locations, and in both related and unrelated contexts. This is a useful feature overall, but it may be inappropriate at certain times. Additionally, it is likely that the timing of these transitions between contexts should follow certain rules and that there will be some degree of nuance in their design. When the same social presence can assist a person in multiple aspects of their life, it is important to understand where social and personal boundaries lie in terms of switching from one physical or topical domain to a completely different one:

RQ4: How, if ever, should re-embodying agents cross contextual boundaries?

Given the futuristic nature of these questions, we utilized structured User Enactments (Davidoff et al., 2007; Zimmerman and Forlizzi, 2017; Odom et al., 2012) to explore how service robots should handle personalization and to attempt to address our four research questions. This methodology has a proven track record for gathering important insights on novel technologies.

3.1 Method

To understand how service robots can employ co-embodiment and re-embodiment to personalize multi-party interactions, we designed a series of User Enactments (UEs) (Davidoff et al., 2007; Zimmerman and Forlizzi, 2017). UEs use low-fidelity prototypes and Wizard-of-Oz methodology to immerse participants in several “possible futures”. By experiencing interactions with mock-ups of future technologies, participants can reflect critically on what they saw, did, and felt, and compare experiences to one another. UEs work especially well in exploratory research, where social mores have not yet emerged, and where there are no existing design patterns. We ran two participants at a time and interviewed them together, which enabled co-discovery and surfacing of knowledge and ideas that one person alone might not have recognized (Lim, Ward, and Benbasat, 1997). Participants signed up together and knew each other, which improved the authenticity of the group experience.

3.1.1 Study setup

The study took place in a lab that was divided into four separate “rooms” by rolling floor-to-ceiling walls. We used scripts that were the result of several weeks of brainstorming and acting out service interactions. The robot was a custom-built exemplar designed for service tasks (see Figure 3.1). The body was made of cardboard with an exterior paper layer. The head was a Kubi desktop telepresence robot with an iPad. We used an iRobot Create as the base. The robot stood about five feet tall and moved at a rate of about half a meter per second. We used Google Cloud Text-To-Speech with five different voices to generate the agents’ scripted speech in advance, and we kept a repository of Google TTS-generated common phrases so that the agents could respond to unplanned deviations. We used three cues to communicate agent identity: each agent had a distinct name, a distinct voice, and a “profile picture” that would appear on the screen whenever that agent was meant to have control of the robot[1]. A researcher controlled the robot and the agents’ voices. The robot followed the same paths each time, so there was minimal variation in its movement. The wizard, who was the same researcher throughout the study, followed a defined script for movements and verbalizations and was instructed to deviate from the script only if the interaction with the participant required an alternative or unique response.

3.1.2 Agent configurations and environments

We designed three agent configurations to explore different interactions that might appear with future service robots (Figure 3.2). We chose these as an initial foray into the design space because they are (1) distinct enough from each other to facilitate critical reflection about ways in which public-facing robots can create a sense of personal connection, (2) conducive to social interaction with multiple people *and* multiple agents, and (3) testable with human dyads (a “single-agent, many-people” configuration limits exploration of certain questions). We utilized a structure that appears similar to a 3x3 study

[1]The software that ran the wizard’s end of the interaction can be found at `https://github.com/AutonomyLab/create_autonomy` and `https://github.com/CMU-TBD/HRI20-Not-Some-Random-Agent-Personalizing-Service-Robot`

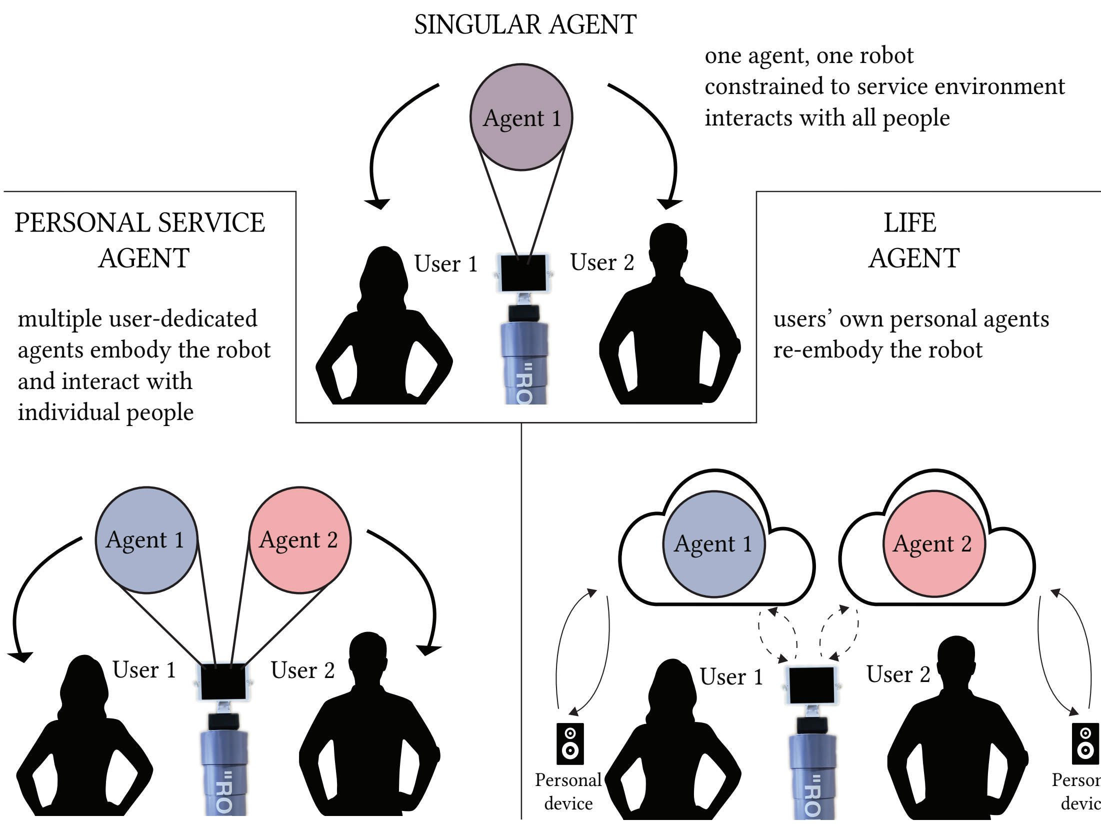

FIGURE 3.2: The three configurations.

design to ensure good coverage of various permutations of contexts and configurations. The added structure helped us cover a vast design space relative to re-embodiment and co-embodiment and avoid overly redundant scenario combinations.

Agents

We iterated on concepts to reach three designs for service robots that personalize interactions.

Singular Agent. This configuration consists of one robot embodied by one agent and is essentially a baseline, i.e., the common paradigm in present-day robots. A Singular Agent (SA) is affiliated with the space(s) it is in and owned and maintained by the service. The agent has information about and "knows" its regular customers. Here, we explored perceptions and impressions of one agent that stores and uses information from multiple repeat users.

Personal Service Agent. A logical step up in service delivery is when a service agent is owned and maintained by the service provider but personalized to each customer. We call this concept a Personal Service Agent (PSA). PSAs are personalized agents assigned and curated by a company or institution. Multiple PSAs can exist in a single physical embodiment. Individual interactions with PSAs are *one agent per user* within a single environment. Because these agents are permanently affiliated with the same service and may need to say the same thing to two concurrent customers, we posited that the PSAs could speak in unison (in a "chorus") to communicate the same message to different people at the same time. With this configuration, we were interested to learn: Should co-embodying PSAs be aware of each other's conversations? How should they talk to each other? We also wanted to explore privacy concerns about agents sharing a data source.

Life Agent. A third option is for each service robot to host multiple individual, personalized AI assistants that are accessed by their users in all aspects of their lives. In the Life Agent (LA) configuration, agents are able to re-embody robots and other devices as needed. Each time the LA re-embodies, it can access the physical capabilities of its current housing and the data specific to the current environment. Thus, it can do tasks with different physical and information demands while allowing the user to interact with any number of unfamiliar devices through the same familiar social intelligence. Pertinent questions are the perceived relationship between LA "software" and robot "hardware" and the evolving social role of this type of integrated AI personal assistant.

Service environments

We designed three environments to examine the influence of service context. These were deliberately chosen to probe issues related to privacy and security, comfort, conversational design, long-term interactions, and social roles. We implemented personalization differently in each environment: in the hotel, it was addressed in terms of food preferences; in the department store, transaction records; in the clinic, medical history.

Quick Care Clinic. Participants entered the clinic together and the robot welcomed them each by name. Then, it guided each participant through the processes of checking in, waiting in a waiting room, and receiving a flu shot. In the LA configuration, P1's agent alerted them that a package had arrived at their home, and P2's agent notified them that an upcoming flight was delayed. The LAs used language that was more

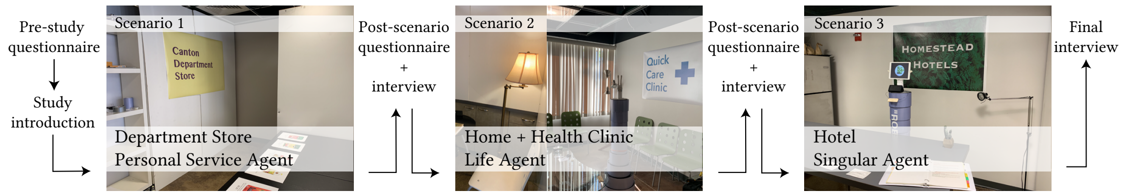

FIGURE 3.3: An example trial from the study. Participants experience the department store with Personal Service Agents first, then the clinic with re-embodying Life Agents (which follow them from home), then the hotel with a Singular Agent.

colloquial (e.g., "Have a seat" vs. "Please sit down") to connote a long-term personal relationship. In the clinic, we sought participants' impressions of agents' ability to use domain expertise and reveal potentially sensitive information.

Canton Department Store. The store environment mimicked two sections of a larger department store. The robot greeted both participants by name, asked (or, for LA, verified) what they were looking for, helped them find the items, and processed payment using a credit card on file. This allowed us to explore how robots should use and talk about personal data in a public space as well as how a robot might handle personalization in a non-personalized environment like a store.

Homestead Inn. In this scenario, we had participants ask a hotel concierge robot for nearby dinner recommendations in an unfamiliar area. Before the interaction, each participant was given a list of dietary, location, and budget requirements, with the goal of finding a restaurant that met both sets of criteria. The agent greeted participants by name and recommended restaurants based on known information about the users and general customer ratings. In the PSA and LA designs, each agent searched for a restaurant on behalf of its own user. Here, we explored how a robot utilizing co-embodiment might engage in a negotiation-like exchange to help users come to a joint decision.

3.1.3 Participants

We recruited 48 participants (24 pairs) via fliers, word of mouth, internet posts, and a local online recruitment tool. Participants were between 20 and 76 years old ($M(SD) = 39.3(17.6)$) and had a variety of personal and professional backgrounds. 25 participants self-identified as female, 21 as male, and 2 as other. They interacted with computers regularly, $M(SD) = 6.48(1.25)$ on a seven-point Likert scale that ranged from never (1) to multiple times per day (7). They interacted with AI assistants less frequently ($M(SD) = 3.31(1.91)$), had some familiarity with robots ($M(SD) = 3.19(1.60)$), and had relatively favorable impressions of robots before the study ($M(SD) = 5.46(1.34)$ for an average of five correlated ($\alpha = .73$) questions about trust and goodwill toward robots). No participants were technical students at our institution.

3.1.4 Study procedure

After consenting to the study, participants filled out a pre-study questionnaire to collect demographics, experience with smartphones and computers, and preexisting associations with robots. A researcher then introduced the study, asking the participants to take on gender-neutral, study-assigned first-names and imagine that they were friends from work. During the introduction, the researcher stated that the goal of the study was to have participants help the team experience and critique potential future interactions with smart technologies. Participants then experienced each service environment with a different agent configuration (three scenarios). We counterbalanced the order of both environment and agent configuration to mitigate the interference of novelty effects in participants' experiences of each of the nine environment-configuration pairings. This meant that 16 participants (8 pairs) experienced each pairing (see Figure 3.3 for an example). We conducted semi-structured interviews with both participants together after each scenario and a final interview at the end of the study. The study took about 90 minutes, and participants were compensated $35 USD each. The protocol was approved by our Institutional Review Board.

3.1.5 Analysis

We identified several hundred meaningful quotes from the interviews, during which participants had an opportunity to respond to questions, react to probes, and reflect freely on their experiences. Our qualitative approach to our data was a thematic analysis in the form of (1) iterative affinity diagramming (Beyer and Holtzblatt, 1997) and (2) application of categorical and sub-categorical labels to quotes based on the clusters that emerged during the affinity diagramming. This approach is used to draw out patterns and themes to explore non-existing, future interactions through UEs (Davidoff et al., 2007; Odom et al., 2012). The analysis was conducted primarily by two authors (one was personally involved in data collection, one was not) who met for multiple hours on several occasions to extract, interpret, and group the data together. They discussed with two other authors after each round of analysis and periodically consulted the remaining authors and a non-author researcher who was less familiar with the details of the scenarios.

We also took special note of responses to three specific questions about (1) acceptance of facial recognition, (2) the *chorus of agents* interaction, and (3) which configurations were most comfortable. We utilized post-scenario questionnaires to assess trust, social attributes (modified from Carpinella et al., 2017 and Bartneck et al., 2009b), and groupness, but results were fairly uniform across agent configurations and service settings. While our approach was primarily bottom-up, we referred back to our guiding research questions to inform the interpretation of the quotes with respect to our research focus.

3.2 Findings

Through iterative analysis of our interview data, we uncovered insights pertaining to our research questions and discovered new themes. We compared a robot embodied by a Singular Agent (baseline configuration) with two variations of co-embodiment: agents owned and managed by the service and agents maintained by the user. Participants generally accepted re-embodiment and co-embodiment, but had some concerns about how re-embodiment might be controlled and how co-embodying agents might exchange data. They did not particularly like PSA, finding the two unique agents to be "redundant" (122B) without adding value. When participants had strong feelings about re-embodiment, these were about the personal nature of LAs. When they had strong feelings about co-embodiment, they were about the concept of multiple software intelligences within one robot. Thus, we report mostly on differences between the LA and SA designs. In the quotes we cite, *Alpha*, *Moon*, *Saturn*, *Basil*, and *Sunflower* refer to the five agents: Alpha is SA; Moon and Saturn are PSA; Basil and Sunflower are LA.

3.2.1 Preference for a Life Agent

Most people (22 participants) thought **a universal Life Agent was the most comfortable design**, followed by a Singular Agent (13 participants), and, finally, a Personal Service Agent (5 participants). Three participants found SA and LA equally comfortable, and 5 had no preference or did not answer the question. In general, participants thought that interacting with a familiar, private agent embodied in public robots would provide a smoother and richer experience. A singular agent was comparable to "just some random person" (119A) that would have neither out-of-context data nor a personal history with the user.

Personality

Participants placed high value on the capability of customization of robot personality and identity attributes. Many wanted robots to exhibit certain character traits when embodied by their own agents, sometimes focusing on traits that would align with or affirm personal values. For example, participant 110A wanted their agent to be hard on them. Participant 101B said, "I want it to be sarcastic because that's how I am. I want it to compliment me. It's like another friend." Some had specific voice characteristics in mind pertaining to gender or dialect: 102B suggested that an agent on the East Coast use East Coast slang, and 101A wanted an agent with a Nigerian or British accent.

Some participants went so far as to say that agents should remind them of their friends or themselves—even to the extent of taking on corresponding voice and speech characteristics. Participant 110B elaborated that a "cool, calm, and collected" person should have a matching robot. This idea is evocative of the well-known finding from sociology that people feel most comfortable socially interacting with people similar to themselves (Lazarsfeld, Merton, et al., 1954; McPherson, Smith-Lovin, and Cook, 2001).

- I'd want it to embody like a personality of my friends, just because you enjoy hanging out with your friends. (107B)
- Though I think it would be creepy, and I probably wouldn't do it, you should [...] have the choice to use your own voice. (103B)

Emotional support

An important function of the LA design is its ability to provide comfort and support. When reflecting on the clinic, several participants mentioned that **in situations that might be stressful or emotional, having a familiar agent would be "comforting"** (125A). Participant 123A mentioned that for someone afraid of shots, their LA should be able to "read that about [them]", and 113A said, "If you're feeling anxious [...], it's nice to have old friendly Basil along who knows everything about you." A few participants thought that robots were more flexible, less distractible, and less likely to get flustered or frustrated than humans; therefore, they were well-suited to jobs requiring patience and calmness. However, most people who alluded to empathy were more of the belief that it is a distinctly human quality that will be difficult or impossible to embed into robots' behavior (e.g., Takayama, Ju, and Nass, 2008). Re-embodiment has potential to augment robots that would otherwise seem impersonal or unsocial with empathetic characteristics just by virtue of feeling familiar and "known" to their users.

Robots can use co-embodiment and re-embodiment to help people feel more comfortable and at ease in unfamiliar spaces, but this raises a set of special design challenges. We discuss the two most significant of these: (1) giving users a sense of control over the interaction and (2) adapting the non-human behaviors of re-embodiment and co-embodiment to human social norms.

3.2.2 Context-crossing and uncertainty concerns

Because co-embodiment was novel, participants were not able to easily anticipate what an agent was going to do next. This became a problem predominantly when LAs had knowledge of participants' personal information, since it was not clear in what (potentially inappropriate) context the agent was going to make use of it in public. Some participants suggested ways to be more in control over interactions with LAs: customizing personality through a questionnaire (116A), using a settings menu to define the nature of the human-agent relationship, or adjusting the LA's conversational style on-the-fly (123B). Many people also felt that **automatic context crossing through re- and co-embodiment should be a toggle setting** such that users could decide, either permanently or for a period of time, to "turn that feature off" (105A).

Control over context crossing

Reactions to the context crossing behavior (i.e., getting non-health-related, robot-initiated personal notifications while at the medical clinic) were mixed. Some participants found this useful, while others thought it strange, awkward, or otherwise an unwelcome social violation. Some expressed surprise when the notification first came in but imagined adapting to such interruptions over time. Some participants noted that an agent that crosses context provides utility by leveraging instantaneous knowledge of remote situations to alert users to information that affects their schedule, safety, or health. For example, it may be appropriate for a user to receive a flight update while at a medical clinic because that can affect their plans for the day. However, inability to anticipate a Life Agent's behavior also led to concern that it might inappropriately surface "out-of-context" information in front of others, oblivious to the incongruous social setting.

Additionally, a universal LA blurs the boundaries between aspects of life that are otherwise separate, and the resulting bleed-through may not always be desirable. For example, 118A said, "There's some universal information like contact lists and stuff like that. But for the most part work should be work and home should be home, should be separate, limited data passing."

3.2.3 Agents are social actors in groups

"Appropriate social behavior" for social agents and robots is not a universal constant: both social context (i.e., the size and composition of the interacting party) and situational context (i.e., the space, place, and task at hand) can change how it should be defined. We found evidence that what is perceived as appropriate social behavior (or lack thereof) of a re-embodied robot may be dramatically impacted by the presence of others. We also found different impressions of social behavior in our three different environments.

Conversational intelligence and social norms

Following conversational norms refers to appropriate physical distance, politeness, common ground, and listening behavior. During interactions among multiple humans *and* a robot, these norms are already at play. Participants felt strongly that a robot should follow norms: 121A said, "Saturn cut me off! [...] If I don't finish, please don't speak!" The field of HCI has long known that people treat technology socially (Nass, Steuer, and Tauber, 1994; Nass, Fogg, and Moon, 1996) and expect agents to have some social intelligence (Nass et al., 1995). However, co-embodied robots encounter special challenges in the way of appropriate conversational behavior. In our study, the coordination of multiple agents sometimes complicated conversational turn-taking, producing "unnatural" (116B) and awkward experiences. Matching or mismatching social norms can also manifest in physical behavior:

- The robot had rolled over to help Alex and then I was still over there and it just turned in my direction and sort of shouted at me instead of coming over to me to talk to me. (122B)

How co-embodied robots handle these norms can also influence or be influenced by morphology. For anthropomorphic robots, in which lifelike physical features reinforce identity, it may be more difficult to communicate the presence of multiple agents.

Understanding existing relationships

Participants believed that it is important for robots to acknowledge an awareness of relationships and history among human members of a group and treat them accordingly. If the humans are strangers, for instance, the robot should "give them their space" (107A). For some, **a robot's ability to exhibit an understanding of human relationships may be a determiner of acceptance of co-embodiment**, especially when interpersonal trust is critical, as in a medical setting.

- How did that agent know that we were even okay getting recognized in each other's presence? (123A)
- If we feel comfortable enough as coworkers to go to the clinic together, I think we can share the same robot body. (119B)

The behavior of agents in a group setting can also influence the way humans perceive and interact with each other—both in the short term as they navigate a conversation, and in the long term as they form lasting impressions of each other. Our interviews suggested that this mediation-like outcome is desirable in low-risk situations that already lend themselves to some degree of casual human social interaction (e.g., the hotel scenario). Through LAs, one person can "learn about the other person you're with very quickly [...] I didn't know he was vegetarian"(107A).

Did it work?

We found that people wanted robots to use human-readable signals to continually communicate information about their status, including multitasking ability, current load, and general capacity. This was prioritized over both efficiency and humanlike social behavior. Even when information did not need to be repeated out loud for an interaction to continue, several participants wished that they had gotten some sort of confirmation that the robot had in fact heard them correctly and performed the task as it claimed it would. This was especially true when accuracy was important and perceived risk was high—e.g., when confirming that it was safe to get a flu shot or that the correct credit card had been used. This is somewhat consistent with prior work, in which people wanted robots to verbally acknowledge the receipt of personal information, even without repeating all of it aloud (Tan et al., 2019).

Who has the floor?

There was a great deal of concern about how co-embodying agents would negotiate multiple users with independent needs and interests. Many people requested that a co-embodied robot provide a "clear indication" (121A) when one agent's interaction ends and another's begins, or when one agent has "handed off" control of the robot to another agent:

- It didn't say, like, Sunflower logging off, Basil logging on, or they didn't switch their icons or it didn't say, like, bye Sunflower, it's Basil's turn now. (109B)

Prior work established that simple movements can go a long way in communicating to users what a virtual agent (Thomas, Johnston, and Thomas, 1995) or robot (Szafir, Mutlu, and Fong, 2015) is about to do. More work is needed to understand how a robot designed to convey multiple "characters" or "personalities" at once could express intent and how the agents embodied in such a robot should negotiate control over that expression.

Inter-agent relationship

There were strong, polarized reactions to the PSAs speaking at the same time. Of our 48 participants, 22 were receptive to the "chorus", 20 were uncomfortable with it, and 6 did not perceive it. Negative responses were rather extreme: participants described the chorus as "an ominous flavor" (109B), "weird" (115B), "creepy and horrible" (122B), and "completely unnerv[ing]" (109A). To better understand these reactions, we affinity diagrammed 31 related quotations. Comments fell into five categories: negative feelings, positive feelings, appreciation of utility, functional complaints, and indifference. Though a few participants were excited about the agents' simultaneous speech, positive feelings mostly took the form of passive acceptance rather than enthusiasm. Many negative feelings stemmed from the fact that it is an extremely non-human behavior.

For both PSA and LA, people did not think the co-embodying agents had a social relationship to each other. They did not believe that the agents would intentionally exchange private information, but they worried that a single robot being embodied by multiple agents could lead to their personal information being "mixed up" (112B) with someone else's due to a mistake or malfunction. People found the idea of inter-agent social conversation creepy and, consistent with prior findings (Luria et al., 2019), feared the prospect of agents "talking behind their back". The exception was negotiation: if agents could coordinate to balance users' preferences or needs, they should. We observed this in the form of overall positive responses to the PSA interaction in the hotel. In other words, **if agents verbally communicate with each other the way humans do, it should only be in immediate service to the user**.

3.2.4 Flexible role conflicts with expertise

We observed **a belief that the more expertise a skill required, the less likely a Life Agent would be to have proficiency in that skill**. As in prior work (Luria et al., 2019), participants had doubts about a "jack of all trades" agent, fearing that it would in fact be a "master of none". In the questionnaire, ratings of trust were lower for the LA, which is intended to serve in multiple domains and embodiments, than for the SA, which is tied to one domain and embodiment and therefore may be more readily considered (and trusted as) an "expert". Beyond this, some participants generally doubted the ability of robots to have real expertise in a non-technological or human-centric domain, or one in which judgment and accuracy in the face of ambiguity are critical (this is similar to (Takayama, Ju, and Nass, 2008). Concerns about expertise were most prominent in the clinic scenario: 9 participants commented on it in the clinic vs. 6 in the store and 2 in the hotel. Participant 119B said that an LA would be trustworthy "if it was a fairly routine problem", but with "a bunch of mystery ailments, I would definitely want a second opinion". Some people commented that upon getting wrong information in a store, "you can find it yourself" (112B), but when it comes to health, e.g., "wrong medicine" (112B), non-experts cannot correct mistakes.

3.2.5 Personal data and privacy

No participants reacted negatively to being recognized upon walking into the clinic. We asked about facial recognition in the clinic setting to explore recognition in the context of private and potentially sensitive information. Even though we did not ask explicitly about it in the post-scenario interviews for the store and hotel, participants took note of it in all three environments. An important characteristic of re-embodiment is that a user's data can move with an agent between robot bodies. This sparked some concern about data leaking from a trusted source to an unknown entity. On the other side of the coin, when an agent was their own, some participants had an increased sense of security—all of their information was concentrated in one place and they did not have to share it in every new context. Instead, a Life Agent could appear and make use of the relevant data. This raises an interesting design challenge: can a robot's behavior indicate that a user's data has left its hardware?

3.2.6 Other findings

A few participants mentioned wanting the robot to have eyes or a face, and some (109A, 111A, 112B) suggested using different modalities (e.g., voice and text input) to ensure that it can be used by older adults and people with disabilities. Many participants did not notice the agents' different voices but noted the changing "profile pictures" and distinct names. As such, voice alone is probably not a strong enough cue to signal agent identity early in a human-agent relationship. Interestingly, this contradicts the original finding from Nass' Computers Are Social Actors experiments (Nass, Steuer, and Tauber, 1994), which found that different voices elicited different social attributions, even in interactions with a novel system. Another theme was societal implications of the futuristic technology we presented. Several participants noted skin tone and accent biases that exist in current face and voice recognition technology. Some expressed concerns about the roles robots will play in the future, including worry that they will not be equipped to carry out the emotional responsibilities humans do and fear that they will take away human jobs.

3.3 Discussion

Our findings address service robot personalization and broader questions about human-robot relationships. Interpreting them requires consideration of the study's limitations: it took place in a lab, agents could not stray too far off-script, the robot was a low fidelity prototype, and only a few people were in the room. Together, these may have contributed to a lack of realism that interfered with participants' ability to fully immerse themselves in the scenarios.

We derive preliminary guidelines for designing the behavior of re-embodying agents, which are of interest to creators of robots and conversational AIs. We also contribute a new way to use UEs to acquire knowledge during an intermediate step of the design process. When a space is largely unexplored, but enough has been learned to spark specific research questions, researchers can add structure (probes, scripts, variations, etc.) to traditional enactments. Thus, they can draw comparisons but leave the experience unconstrained enough to facilitate revelation of "unknown unknowns".

We inquired as to how re-embodying agents should perform their service with multiple users (**RQ1**). We found that participants prioritized social competence and personalization during group interactions. We noted a distinction between personalization of social features and personalization of personal information. Participants in our study envisioned a Life Agent to be able to prioritize information that was specific and pertinent to them (perhaps in contrast to other users) and to build on and draw from that knowledge over the long term, regardless of whether or not its personality and social behaviors were customized. This increased their feelings of comfort interacting with the agent (**RQ3**) and made it generally desirable.

RQ2 concerned the overarching impact of co-embodiment on perceptions of social robots. Co-embodiment was received as (1) *necessarily* concerned with social signaling, and (2) appropriate for friends, but not for strangers. We draw from this two concrete design guidelines for co-embodying and co-embodyable systems. The first is opt-in co-embodiment: robots in public settings can enable co-embodiment, but should not be embodied by two agents at the same time by default; and they should be explicit on whether a third party can gain access to the data from an interaction. The second is clear indications from robots about what ("who") is in charge. Repetitive signaling regarding which agents are being accessed and which users are being helped is critically important for users to understand how to interact with a co-embodyable robot, at least in early interactions.

In response to our research question on contextual boundaries (**RQ4**), we find tension between comfort and expertise: people have difficulty with the idea of one social agent that claims to be equally adroit in all possible domains and embodiments. At the same time, they want to interact with novel robots through a Life Agent that aligns with personal identity and values. This presents a challenge of balancing quality and quantity. We conjecture that embodiment of personal agents in non-personal robots is best used for tasks that are perceived to be relatively low-risk—for example, helping people navigate a building using familiar language or making recommendations in a grocery store based on knowledge about cooking habits. In contrast, when perceived risk is high, as in a medical setting, robots need to prioritize the communication of their expertise over personal connection and emotional support. One approach to mitigating this tension might be to design an agent that communicates that it is acquiring expertise. For example, a Life Agent, upon entering a healthcare facility, might communicate that it is acquiring new expertise in support of the user's interactions with the service. But in some cases, re-embodiment of a Life Agent into a domain-specific robot may be best foregone entirely in favor of clear assurance that a robot is well-versed in the task and solely dedicated to it.

When agents *do* transition across contexts, our data suggests they should clearly express the features that constitute their identity. Defining the minimum cues necessary for users to recognize an agent is a critical part of designing re-embodyable systems. Confusion about how and when re-embodiment has occurred may be tied to discomfort with the concept and, in turn, result in lower acceptance. We used three attributes to communicate an agent's identity: image, voice, and name. In our study, image was a much stronger cue than voice or name. Of course, it is not feasible to take this as an absolute because many robots do not afford projecting an image onto a screen and because visual-only channels make robots less accessible. What we can conclude is that whenever possible, designers of re-embodyable robots should provide a means of visually indicating the presence of different agents.

Finally, our study provokes examination of and reflection on the role of robots in society. The lack of concern with facial recognition by robots in both private and public spaces likely requires a more nuanced inquiry than our study provided. The broader privacy issue of the conflicting interests of multiple stakeholders has a close and complicated relationship with feelings about facial recognition: for example, a few participants felt that facial recognition in a clinic setting would be useful or even necessary for trust, but that in a commercial setting, it would be in the interests of the company rather than their own and an inappropriate violation of their privacy. Future research into where and when it is acceptable for robots to use facial recognition, and how storage and usage of that data should be communicated, will benefit the design of service robots from a user experience perspective as well as an ethical one.

The preference for a customizable Life Agent similar to oneself raises questions about defaulting to designs that reinforce people's tendency to gravitate towards similar others. The non-human characteristics and customizable capabilities of social AIs and robots may make them conducive to designs that challenge social biases rather than conform to them. Some participants asked to have a personal agent with qualities that complemented, rather than matched, their own. This gives credence to the idea that while people value the familiarity and support of agents that are like themselves, they may also accept, and even desire, dissimilar agents (Isbister and Nass, 2000).

3.4 Summary and contributions

This study investigated how future service robots can use personalization to interact with multiple users. Through structured user enactments and interviews, we found that people are receptive to the idea of robots that leverage personal information if the user has control over the information. We also discovered that service robots embodied by multiple agents can make people more comfortable with group interactions by demonstrating an understanding of pre-existing human relationships within the group. Our work sheds light on the role of flexible agent embodiment during interactions with service robots, and suggests design guidelines and directions for future research on the topics of re-embodiment, co-embodiment, and personal human-robot interactions that occur in public.

The work described in this chapter makes the following contributions:

- We identified two possible configurations for re-embodying and co-embodying agents: "Personal Service Agents" and "Life Agents".
- We found that people generally preferred to interact with personalizing agent identities that re-embody across services over service-specific personalizing agent identities and embodiment-specific identities.
- Our participants' comments revealed affordances of co-embodying and personalizing agents that would provide value: emotional support and personality customization.
- We also identified concerns surrounding such agent behaviors, and possible ways to assuage those concerns:
  - People may worry about uncontrolled context-crossing of agent identities; therefore, this should be a toggle setting that users can control.
  - Lack of understanding of social context can bring about perceived and real personal privacy risks and awkwardness; therefore, co-embodying agent identities should follow social norms, legibly signal when they are accessing different people's data and directing interaction to different users, and communicate their understanding of social context (i.e., the relationships among multiple simultaneous users).
- We found that people are uncomfortable with service agents that communicate with each other in humanlike ways when not directly responding to a user's request.
- We built a custom service robot that can be used in and modified for future service design research, as well as inspire the design of commercial service robots.

- We pioneered a new variation on the method of user enactments that emphasizes adding overall structure, comparisons, and events in order to better understand intermediate-level design knowledge (this is elaborated in Chapter 9).

Chapter 4

Comparing Personalized Interactions with Fluidly-Embodied Service Robots: Storyboards Study

Much of this chapter was previously published as the scientific article:

Samantha Reig, et al. (June 2021.) "Social Robots in Service Contexts: Exploring the Rewards and Risks of Personalization and Re-Embodiment". In *Designing Interactive Systems Conference 2021 (DIS '21)*.

Permission to reproduce the text as part of this dissertation is included in the original publication's copyright.

This study builds on the work discussed in Chapter 3 through an online *speed dating with storyboards* study. We set out to better understand when agent identities embodying social robots in front-line service roles should act more like human service workers, and when they should take advantage of capabilities that would be difficult or impossible for people (e.g., instantaneously accessing customer records, enacting many different roles, or re-embodying). Because they are complicated to set up and execute, UE studies work with small numbers of participants. In order to learn in more depth how agents might use co-embodiment and personalization to mediate relationships among an individual, multiple embodiments, and a service, we proceeded with this line of work employing a method that allowed us to investigate whether the preferences observed from small samples might generalize by using a larger audience.

We conducted an online study using storyboards (Truong, Hayes, and Abowd, 2006) that described common service encounters that many people would find familiar. However, our storyboards featured robots in roles that are typically populated by human service agents. Using familiar situations helps participants reflect on the future based on their felt-experience of the present. Our use of storyboards to communicate these familiar yet future experiences allowed us to explore the two selected concepts (customer identification and robot re-embodiment) across several service contexts in a single study. Distributing the storyboards via an online study recruitment platform enabled us to rapidly determine whether insights from UE studies conducted with small numbers of participants could scale, and whether there are differences in the appropriateness of robot behaviors across different service contexts.

The study makes two novel contributions. First, it provides a more nuanced understanding of people's beliefs about appropriate robot behavior and boundaries for service robots. Specifically, we illustrate how service settings where people are expected to make a personal appointment (as with a hair stylist or physician) evoke different notions of what constitutes appropriate and acceptable robot customer service than do settings where personal appointments are less common. We also describe qualitatively-derived possible impacts of perceived personal risk and perceived similarity of tasks on the acceptability of agent re-embodiment. Second, our work provides one example of how design researchers might advance insights that emerge from UE studies or other small-scale, qualitative research studies. We show that a relatively large N online study that uses storyboards and questionnaires with free response fields can deepen the knowledge gained from early exploratory work on novel concepts. Researchers seeking to probe issues and questions that arise at the intermediate stage of knowledge-gathering (see Höök and Löwgren, 2012; Löwgren, 2013) on a variety of topics can draw on this method to advance understanding.

4.1 Claims

We sought to dive more deeply into findings that were raised in our two previous studies. In an iterative ideation process, we narrowed down to two prior aspects of personalized service interactions: user recognition and robot re-embodiment. We set out to better understand *claims* that we derived from the literature on both aspects, which suggested that people's reactions are strongly influenced by the service context. We emphasize that our proposal and analysis of these claims is purposefully *exploratory*. Rather than assert a prediction and then use empirical methods to test it, we use the claims to guide an evolving understanding of intermediate-level knowledge (see Höök and Löwgren, 2012; Löwgren, 2013) of design concepts.

For personalization, we focused on two findings regarding people *being identified* by robots. We extracted two claims to explore:

- **C1-Pers:** People will be bothered when a robot identifies them in a service where they would not expect to be identified.
- **C2-Pers:** People will be less bothered by being identified when a robot uses customer profile data to deliver something of *value*.

With respect to robot *re-embodiment*:

- **C3-Re:** People will not want robots to re-embody when it involves a large change in social role or expertise.

4.2 Research Approach

Designing the behavior of social robots working within brick and mortar services requires many choices, and each decision likely impacts a customer's holistic experience. Given this nearly unbounded space of investigation, design methods offer an effective approach for gaining insights. Human-computer interaction (HCI) research notes the tension between scientific research that seeks to use complex instruments to exert control over phenomena, and design work that gleans knowledge from complexity through the use of simple tools (Stolterman, 2008). Design methods allow researchers to rapidly explore a broad set of design choices and future situations. This type of exploration is less accessible with hypothesis-driven research, in which many aspects of a design need to be carefully controlled in order to generate new knowledge. We used the method of Speed Dating with storyboards (Davidoff et al., 2007). The mid-level fidelity of *storyboards* allows researchers to rapidly iterate pilots to progress towards stimuli that effectively probe a study's research goals. They also allow participants to experience small sips of many different situations, which helps them gain higher-level insights on what they actually want and expect (Zimmerman and Forlizzi, 2017). This method is useful for drawing participants' attention to specific examples of technology behaviors, especially ones that are set in the future or that do not exist yet (Branham, Wahid, and McCrickard, 2007; Luria et al., 2020; Truong, Hayes, and Abowd, 2006).

4.2.1 Creating storyboards

Over several months, 7 researchers iteratively generated scenarios and then storyboards for a diverse set of service contexts. We structured our brainstorming around situations that were socially complex. Many HRI and storyboard studies focus on one-to-one interactions—but service interactions involve groups, interpersonal relationships, and public spaces. Similar to Luria et al. (2019) and the work discussed in the previous chapter (see Reig et al., 2020), we aimed to produce scenarios that were novel and clearly set in the future, yet realistic enough that participants could readily imagine them occurring in their lives. We initially brainstormed 16 scenarios. We then refined these through four rounds of piloting involving 310 participants. We narrowed down our set of claims based on how the storyboards were received. After each pilot, we retained the scenarios that had the most traction and were most conducive to exploring design nuances in different service settings. Through multiple rounds of piloting, we gleaned insight into which service settings were sparking the most reflection from participants, and which storyboard pairs generated the most telling comparisons and contrasts in the data. We relied on our collective judgment and drew on our evolving knowledge of the design space and research method as we downselected to the final set of storyboards. Piloting also enabled us to identify anything that was confusing (e.g., overly complex robot dialogue) or misinterpreted (e.g., interactions we expected to deliver value but where participants perceived no value). Piloting resulted in 11 final storyboards related to our claims: 4 that addressed C1-Pers, 3 that addressed C2-Pers, and 4 that addressed C3-Re.

Final storyboards

We created two versions for each final storyboard; one capturing the assumed preference according to the claim, and one pushing against it. Figure 4.1 provides examples of storyboard pairs addressing C1-Pers and C3-Re. In the C1-Pers example pair, Pat, a customer, is recognized (1) at an auto shop (where a customer might expect to be recognized) and (2) at a Carnival (where a customer might expect not to be recognized). In the C3-Re example pair, Jerry is at the airport and encounters Bob, a gate agent, who re-embodies into (1) the seat-back entertainment system on the plane, or (2) the flight's co-pilot. Each pair used a unique name for the customer and for the robot, so participants would never view a new storyboard as a continuation of a prior one. We attempted to make the pairs as similar as possible, varying only the features we wanted to compare.

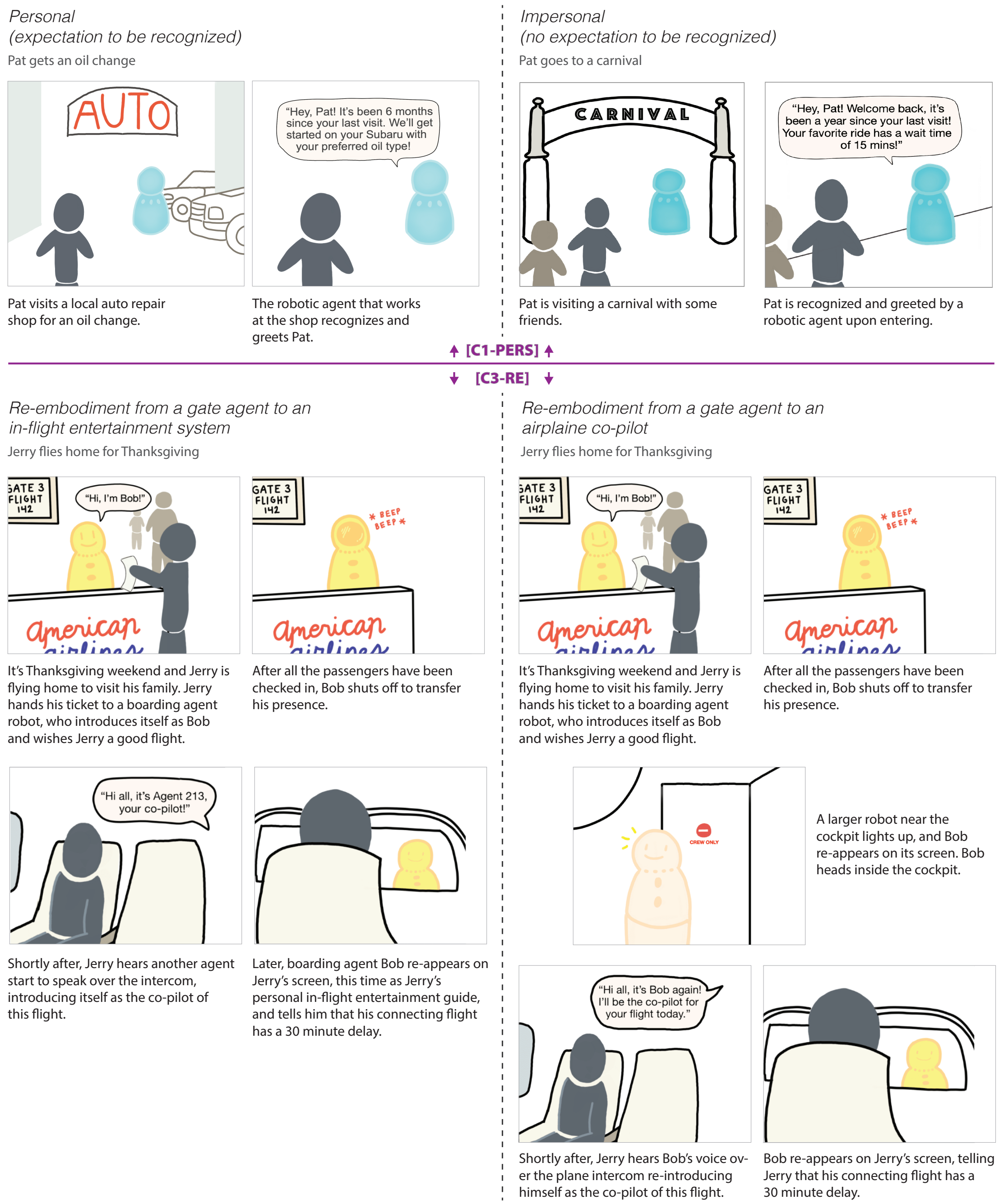

FIGURE 4.1: An example of two storyboard pairs that were compared to address *C1-Pers* and *C3-Re*. Top left: Pat is identified by a customer service robot at an auto shop. Top right: Pat is identified by a robotic employee at a carnival. Bottom left: Jerry interacts with an airline gate agent robot that re-embodies to function as the plane's in-flight entertainment system. Bottom right: Jerry interacts with an airline gate agent robot that re-embodies to function as the plane's co-pilot.

We also purposely kept the narratives fairly straightforward and simple. Research on the fidelity of prototypes (e.g., Davidoff et al., 2007; Buchenau and Suri, 2000; Truong, Hayes, and Abowd, 2006) details the need to choose the "right fidelity" to keep a focus on the phenomena of concern and has shown that storyboards allow for more breadth of search in exploratory research. As stimuli become more specific, the question changes from "Would people want a robot to exhibit this behavior?" to "Would people want *this* robot to exhibit this behavior?" Our storyboards were simple to strike a balance: examples had to be specific enough to give participants something to reflect on, but not so specified that they would overly constrain interpretations and reflections.

Customers were depicted in a consistent, 2D visual style that de-emphasized gender and racial cues, allowing participants to more easily envision themselves in the depicted situation. We drew robots of different forms, sizes, and colors for different tasks and contexts. We also used gender neutral names for robots to reduce any gender effects, as seen in prior work (Tay, Jung, and Park, 2014)[1].

C1-Pers: The first set of storyboards probed customer identification in impersonal vs. personal service contexts. Each storyboard pair included a more personal (e.g., a hair salon) and a less personal (e.g., a department store) context. The customer was identified in each. The paired storyboards (where the setting was the variable we manipulated) were: 1) an office supplies store (impersonal)/a hair salon (personal), 2) a carnival (impersonal)/an auto shop (personal), 3) a department store (impersonal)/a gym (personal), and 4) a grocery store (impersonal)/a doctor's office (personal).

C2-Pers: The second set of storyboards examined if delivering something of *value* mitigated the perceived *creepiness* of being identified. In these storyboards, a robot would share information that made it clear that customer behavior was being observed over repeated interactions with the service. The storyboards were identical in each pair; however, the *value* version had some form of value (e.g., a coupon) following the suggestion that the customer was being tracked. The storyboard settings (each with a value/no value pair) were 1) a fast food chain, 2) a movie theater, and 3) a superstore.

C3-Re: The third set centered on robot re-embodiment and differing expertise. Prior work (Luria et al., 2019) found that people may be concerned about a re-embodying robot having the expertise required to do different jobs. In this work, we used specific examples of contexts requiring different kinds of expertise to examine *why* this might be the case. We probed at whether the "social status" or "prestige" that people associate with different jobs might play into these concerns. We also explored whether the similarity of the roles assumed by a re-embodying robot might impact concerns about expertise. In these storyboard pairs, we varied the similarity of the roles that the robot played before the re-embodiment and after the re-embodiment. The storyboard settings (each with two different versions of the "target" of the re-embodiment) were: 1) a hotel, 2) air travel, 3) physical therapy and massage, and 4) a dentist's office.

4.2.2 Participants

We recruited 204 participants through the online survey research platform Prolific. To be included, participants had to (1) be at least 18 years of age, (2) be fluent in English, (3) have previously completed at least 50 submissions on Prolific, and (4) have at least

[1]See the Supplementary Material of the published paper for the full set of storyboards and questionnaires.

a 90% approval rate for previous submissions. Participants were paid $5.00 USD each. Our study employed attention checks, and 197 participants answered these questions correctly. Most participants (107, 54%) were between the ages of 18-30. Sixty-nine (35%) were 30-45, and twenty (10%) were 45 or older.[2] One did not report their age. Ninety-nine (50%) identified as female, 94 (48%) identified as male, 1 identified as non-binary, 2 self-described as genders not represented in our multiple choice options, and 1 chose not to disclose their gender. The study was approved by our university's Institutional Review Board.

4.2.3 Procedure

After being redirected from Prolific to our survey and giving informed consent, participants answered several questions concerning their prior experience with service settings the study would involve. These questions were intended to control for differences in familiarity and unfamiliarity with the contexts. They then viewed various storyboards exhibiting versions of the behaviors we wanted to explore. Each scenario was presented with a set of closed-ended Likert-type questions on thoughts, feelings, and perceptions about the storyboards and the events within them. Many of these questions were followed by prompts asking for explanations, which were used in our qualitative analyses. Participants took between 10 and 60 minutes to complete the study.

We divided the full set of *paired* storyboards into two groups. We then divided each of those groups into two subgroups to separate each pair of storyboards (one of each pair went into each subgroup). Each participant was randomly assigned to one of these four groups. Three additional storyboard pairs that were related to a fourth claim not discussed in this paper were also included in the Qualtrics survey. This meant that each participant viewed 7 storyboards, and 51 participants viewed each storyboard.

4.2.4 Measures

For storyboards that focused on C1 (personal and impersonal contexts) and C2 (added value), we assessed perceptions of the robot as *creepy* and perceptions of the encounter as *friendly*[3]. For the storyboards that focused on C3 (re-embodiment), we assessed perceptions of the level of *prestige* and *expertise* of each role performed by the robot in the story, perceptions of the robot's *competence*, belief that the service did a good job of creating an agent capable of *multitasking*, whether or not it was appropriate for the agent to *serve in both roles*, and whether or not the agent *should take on multiple responsibilities*. For all scenarios, we asked about perceptions of the encounter as an *improvement over the typical service experience*. Finally, for each scenario, an open-ended question asked participants to explain their ratings for the main variable of interest. For the C1 and C2 storyboards, this question pertained to ratings for how *creepy* the agent's behavior was. For the C3 storyboards, it pertained to ratings of whether the agent should or *should*

[2] Our response categories forced any participant who was 30 years old or 45 years old to choose between two overlapping descriptions. Unfortunately, this error was not caught until after data collection.

[3] "Creepy" and "friendly" are not opposites, but they are concepts with opposing sentiments that could each possibly describe the way it feels to be spoken to in a very personal manner, especially by a robot. Our choice to use these particular positive-sentiment and negative-sentiment words was inspired by the themes from Chapter 3 (see Reig et al., 2020 for the publication reference). Perceptions of creepiness and friendliness were measured via two separate questions; we do not assume them to be mutually exclusive.

not take on multiple responsibilities. Because this is exploratory design research intended to garner direct feedback from participants, we designed the questionnaire items to directly ask about the concepts and perceptions we were interested to understand. Each question was analyzed as a single item. Figure 4.2 shows the response distributions for these variables.

4.2.5 Analysis

To analyze the qualitative data, we took a two-pronged approach. We first looked at the short-answer explanations associated with ratings at the extremes of the Likert items to check for extreme response bias (see Greenleaf, 1992). Our review of this subset of the data suggested that people did indeed have strong positive and negative reactions to the scenarios. Two members of the research team then went back through the full data set, manually reading each response and annotating the findings. In doing so, they made note of which responses were associated with which storyboards, and where the participant's corresponding scale ratings fell (e.g., an individual response might be annotated with "superstore-no value, perceived creepiness=very high").[4] This allowed us to interpret patterns in light of the comparisons we intended to draw between the different storyboard versions. Multiple research team members reviewed these notes and discussed key themes and insights in the data, leading to the insights we discuss in the Findings section and our design recommendations.

We also analyzed the closed-ended questions (Likert-type items) using Welch unequal variances t-tests. We use the results of these analyses to support and help describe the qualitative findings.

4.3 Findings

We asked participants to report on their prior experience in each of the less-commonplace contexts. Most participants had experience with these: 172 had flown on an airplane, 171 had worked out at a gym, 106 had been to a salon, and 152 had taken a car in for repairs. One participant did not have experience with any of these situations.

4.3.1 Impersonal and personal settings

Differences in participant responses revealed nuances regarding the appropriateness of recognizing users in different service contexts. Our previous User Enactments study found that people did not like service robots to recognize them in settings they viewed as impersonal (e.g., a department store), but desired it in settings where they expected personalized service (e.g., a doctor's office). This draws a distinction between two types of contexts: those in which the professional relationship between the service providers and the customer involves a degree of more-intimate interaction, and those in which it maintains more distance. Our findings suggest that **expectation to be (and appropriateness of being) identified by robots is a complex issue** that cannot be reduced to a

[4] We make statements about "participants who found the robot to be creepy" and similar generalizations throughout the Findings section. Such classification of participants is based on their scale ratings; i.e., whether their creepiness scores were lower or higher than zero.

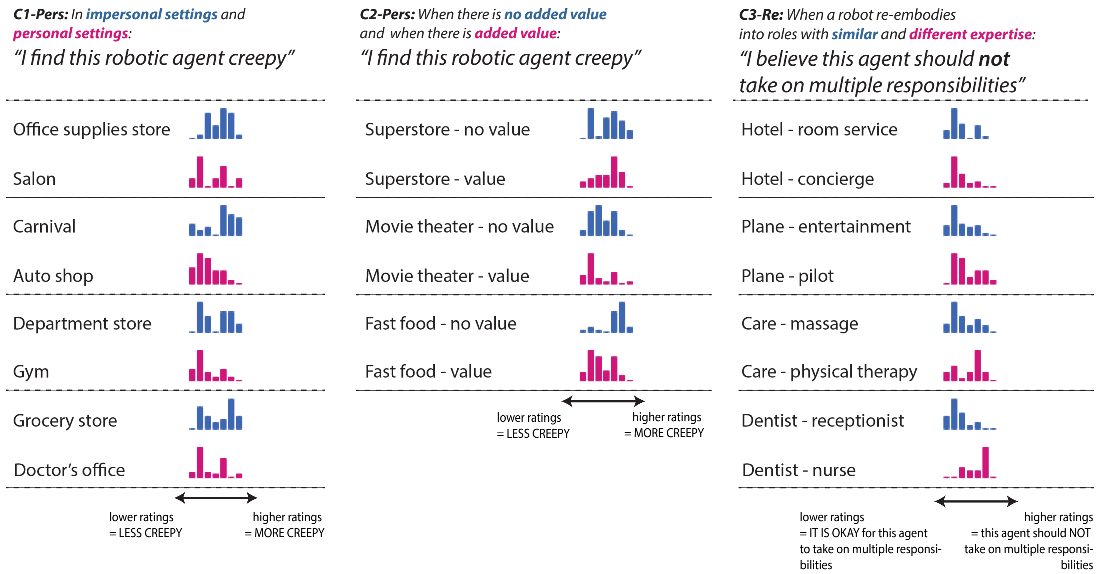

FIGURE 4.2: The distribution of ratings of *creepiness* (C1 and C2) and the *belief that a robot should not re-embody into roles with different expertise* (C3) for each of the 22 storyboards.

simple binary. Overall, being identified by a robot was perceived as less creepy in settings where customers expected to interact with the same service agent, and expected service to be personalized to their individual needs.

Across all C1-Pers scenarios, participants who found the agent to be creepy generally expressed one of three concerns. One, they opposed facial recognition. Many (n=20) commented explicitly on their **discomfort with facial recognition software being employed in a service setting**. None of the storyboards detailed how the robot recognized the customer—rather, participants inferred that it was using facial recognition (in fact, one participant, P29, commented that the robot was *not* creepy because they assumed it to *not* be using facial recognition). Two, participants did not want to be profiled. They explained that the robot's verbal disclosure of the amount of time since the customer's last visit showed the service collected an **unnecessary and uncomfortable amount of information**. (Participants who mentioned this included several of those who explicitly pointed to facial recognition as a concern and several others.) A robot's intention to be friendly or helpful was not enough to justify the profiling. Instead, participants came to their own conclusions about whether the service had reason enough to collect and use personal data. Three, participants shared that being identified by the robot was creepy because it **lacked the human-like characteristics** that could make this kind of interaction seem empathetic. This reaction was similar to findings from our prior work.

Participants who found the personal identification (which happened in all C1-Pers storyboards) less creepy commonly mentioned that it was a **friendly behavior that would make them feel welcome** (e.g., "I think it makes a more personalized experience and makes people feel more welcomed and seen,"—P106), or that it added value to the service experience (e.g., "I love being welcomed. A little compliment goes a long way for me,"—P140). Several participants thought that the robot's behavior was perfectly acceptable because it was no different from how a human in that position would

Storyboard version	Personal or Impersonal	Creepy	Better than typical experience	Friendly
OFFICE STORE	Impersonal	**0.27 (1.76)**	0.08 (1.68)	**0.79 (1.49)**
SALON	Personal	-0.20 (2.00)	**0.15 (1.80)**	0.67 (1.73)
	p-value	0.227	0.832	0.713
CARNIVAL	Impersonal	**0.50 (2.02)**	0.28 (1.91)	0.57 (1.83)
AUTO SHOP	Personal	-0.92 (1.56)	**1.25 (1.38)**	**1.60 (1.14)**
	p-value	**0.0002***	**0.006***	**0.002***
DEPT. STORE	Impersonal	**0.10 (1.88)**	0.56 (1.70)	1.18 (1.47)
GYM	Personal	-1.08 (1.64)	**1.02 (1.36)**	**1.59 (1.12)**
	p-value	**0.001***	0.141	0.119
GROCERY STORE	Impersonal	**0.30 (1.95)**	0.10 (1.69)	0.88 (1.44)
DOCTOR'S OFFICE	Personal	-0.53 (1.83)	**0.45 (1.57)**	**1.31 (1.40)**
	p-value	0.031*	0.290	0.139

TABLE 4.1: M (SD) for each storyboard in the *Impersonal (I) vs. Personal (P)* set (*C1-Pers*). Ratings were on a scale from -3 (strong disagreement) to 3 (strong agreement). Values in bold with * are $< .05$. The higher value of each measure for each storyboard pair is in bold text.

behave (9 responses reflected this sentiment directly, and several others alluded to it). Interestingly, a couple of participants believed that a robot could do the human-facing, empathy-requiring aspects of the job *better* than a human could, and therefore, would prefer robots in service roles involving personal identification (P17 said that robots would not "disrespect and judge" as humans do).

Out-of-place identification.

The carnival/auto shop storyboard pairing demonstrated the largest differential in perceptions of the agent as creepy. Perceived creepiness was higher in the carnival than in the auto shop, the auto shop was perceived to be a better improvement over the typical service experience, and the auto shop encounter was perceived as more friendly (see Table 4.1).

We infer that the large difference between the carnival and auto shop storyboards stems from the fact that **a carnival is at the extreme low end of the *expectation to be identified* spectrum**. For most, a carnival is a novelty event that is not available year-round. While some ride operators and ticket salespeople may be locals hired for a single gig, long-term employees or robotic service workers would likely travel. To be identified by them would be an anomaly. Here, when participants thought that the encounter was *not* creepy, they noted that it was "wholesome and harmless" (P13), and that learning the wait time for a favorite ride was useful. When they were bothered by it, they **commented on the identification being out-of-place and unnecessary**: P45 said, "It would be strange to be recognized personally at an amusement park after a year and for them to know your favorite ride." At $M = 0.5$, carnival creepiness was rated higher than any other storyboard (second-highest was the grocery store, $M = 0.3$). In contrast, participants commented that **in an auto shop, having data on a customer and their visit history is directly related to the service being provided** (e.g., "I'm assuming the robotic agent just has documentation and a log of all its customers and their past services"–P130).

Identification that is necessary for the service.

The doctor's office was at the opposite end of the spectrum. In medical settings, correct and reliable identification is mission-critical. Participants' explanations of their creepiness ratings spoke to this: P22 said, "Doctor offices keep files [...] it is not top secret information. It is reaffirming your identity and confirming," and P175 said, "At the doctor's office I am more comfortable with an AI that is aware of my visit frequency." A hair salon and gym are both somewhere in the middle. In these settings, it is fairly common to be recognized by workers, and in certain cases—for example, when getting a haircut from one's favorite stylist, or when working with a personal trainer—not being identified correctly could be cause for alarm and/or be a detrimental experience. In these two contexts, participants who did not find identification creepy said that it was a useful feature (e.g., "It can help save time", said P44 about the hair salon), that it made sense given the setting (e.g., "Their job is to remember things like this for the experience," said P73 about the gym), and that it did not overstep an interpersonal boundary (e.g., "Not threatening in any way," said P34 about the hair salon). This distinction bore out in the Likert ratings as well: creepiness was higher in the grocery store than in the doctor's office, and higher in the department store than in the gym (see Table 4.1).

4.3.2 When identification adds value

The second concept we were interested in was whether perceived *value* in a service encounter impacts what is acceptable robot behavior. We included a validity check question to determine whether or not participants thought the robot's activity provided some type of value for each scenario. In all three storyboard comparisons, the robot in the "value" storyboard version was perceived as providing value. All differences were significant at $p < .05$.

Overt tracking.

In each of the three storyboard pairs, the pattern was the same: First, **the *no value* storyboard was creepier** than the *value* storyboard. Second, the *value* story was a larger improvement over the typical service experience, and the robot's behavior was more appropriate. Finally, people were more suspicious of robots in *no value* storyboards. However, not all of these differences were statistically significant (see Table 4.2).

In both versions of the movie theater storyboard, identification and user profiling were generally perceived neutrally. Participants noted that a comment from the robot about visit frequency was "an innocent observation" (P75) that was harmless, pleasant, and relevant (e.g., "It's nice to be remembered and recognized as a fan"–P196). In the fast food scenario, many participants who thought that identification was creepy were concerned less with the data collection itself than with the **visibility and obviousness of the data collection**: P45 said, "People don't like to be reminded of how much information businesses and corporations gather about them," and P136 said, "Nobody wants to know how much fast food they've been eating." In a superstore scenario where a robot asked a customer if she had recently had a baby, many people mentioned that **the tracking required to make such a personal inference was unnecessary and over the threshold of what was comfortable and valuable**. For example, P141 said, "It knows about her baby and personal life and that's weird," and P54 said, "It is not appropriate for a robot to ask such personal questions."

Storyboard version	Value or No value	Creepy	Better than typical experience	Acts appropriately	Suspicious
	No value	**0.18 (1.78)**	0.16 (1.57)	0.42 (1.75)	**-0.26 (1.83)**
SUPERSTORE	Value	0.08 (1.75)	**0.76 (1.44)**	**0.76 (1.63)**	-0.43 (1.93)
	p-value	0.782	0.052	0.326	0.656
	No value	**-0.41 (1.61)**	0.16 (1.36)	1.10 (1.42)	**-0.94 (1.59)**
THEATER	Value	-1.0 (1.73)	**1.38 (1.65)**	**1.66 (1.39)**	-0.94 (1.90)
	p-value	0.080	**0.0001***	0.051	0.997
	No value	**0.61 (1.86)**	-0.72 (1.54)	0.15 (1.51)	**0.24 (1.75)**
FAST FOOD	Value	-0.52 (1.65)	**1.10 (1.40)**	**1.48 (1.23)**	-0.77 (1.88)
	p-value	**0.002***	**<.0001***	**<.0001***	**0.007***

TABLE 4.2: M (SD) for each storyboard in the *Added Value* set (*C2-Pers*). Values in bold with * are < .05. The higher value of each measure for each storyboard pair is in bold text.

Social privacy violations

A theme that emerged in C1 and C2 storyboards was concern about the possibility of social privacy violations. Participants were worried that in public settings, regardless of recognition by a robot or value of the service, **strangers might overhear conversations between a customer and the robot, which could affect their trust in the robot and in the service**. This sentiment was particularly strong in the superstore scenario, where the robot said aloud that the customer was probably looking for diapers. Participants did not like the idea of this private information being made known to anyone within earshot, and thought it could even "threaten a customer's safety" (P187). There were no significant differences in any of the ratings of the service encounter as creepy, better than the typical service experience, appropriate, or suspicious between the *value* and *no value* versions of this storyboard. It is likely that the public announcement of private information (which occurred in both storyboards) was so noticeable and so unappealing to many participants that it undermined their likelihood of caring about or even noticing the value-related difference between the two.

Taken together, the findings related to *C1-Pers* and *C2-Pers* suggest that being welcoming and friendly—and even concretely helpful—is not reason enough for robots to recognize and profile customers in most settings. Most customers will only respond positively to this behavior in scenarios where a *failure* to correctly confirm their identity would either be genuinely worrisome (as in a doctor's office) or seen as poor customer service (as in a hair salon where customers book appointments ahead of time). Essentially, service robots should identify customers where their human counterparts would be likely to do so as part of the service rather than as a personal quirk or as a result of repeated interaction—and likely not anywhere else.

4.3.3 Re-embodiment and expertise

Multiple roles: why not?

In general, **participants seemed comfortable with the idea of robots taking on multiple roles requiring different kinds of expertise**. In all eight of the C3 storyboards, average ratings for the statement "the agent should *not* take on multiple responsibilities" were relatively low (see Table 4.3). Justifications of these ratings evoked the convenience of streamlining the interaction (i.e., having one party keep track of all of the information across multiple touchpoints), good faith in a robot's ability to handle multiple tasks at once better than a human, and a sense of indifference: "Why not?" (P41, P56, P72, P96, and others).

Initially, we suspected that concern with re-embodiment might be associated with the prestige differential between two roles. But as with C1, we discovered more nuance. Perceptions about individual storyboards also differed in ways that revealed patterns associated with specific contexts and service roles.

Hotel: a single service in a single domain.

In one version of this storyboard, people were comfortable with a robot serving in two roles (hotel maintenance and room service), even though they perceived a relatively large difference in expertise. Most (26 out of 33) explanations for ratings suggesting positive perceptions of this storyboard reflected an assumption that the two roles could be executed well enough by the same robot: P52 said, "Both tasks require low to medium maintenance skills and I assume that configuring the robot to carry out both tasks should have minimal side effects," and P31 said, "It seems natural for a single robot to do these tasks." People were also generally comfortable with a robot serving in the maintenance and concierge roles despite a nontrivial difference in perceived prestige. For this scenario, we found no significant differences for any of the Likert items.

Context	Second role	Better than typical experience	Should NOT take on multiple responsibilities	Should NOT serve in both of these roles	Competent
HOTEL	Food	0.84 (1.23)	**-1.06 (1.52)**	**-0.94 (1.64)**	**1.92 (0.81)**
	Concierge	**0.92 (1.19)**	-1.14 (1.40)	-1.16 (1.36)	1.68 (1.06)
	p-value	0.733	0.789	0.467	0.211
AIR TRAVEL	Seat-back	**0.62 (1.24)**	-0.88 (1.67)	-1.10 (1.61)	**1.54 (1.18)**
	Pilot	-0.04 (1.38)	**-0.22 (1.79)**	**0.00 (1.90)**	1.45 (1.28)
	p-value	**0.014***	0.063	**0.003***	0.714
PERSONAL CARE	Massage	**0.50 (1.57)**	-0.69 (1.71)	-0.62 (1.83)	**0.98 (1.37)**
	Phys. therapy	-0.09 (1.85)	**-0.09 (1.85)**	**0.83 (1.34)**	0.02 (1.67)
	p-value	0.096	0.097	**<.0001***	**0.003***
DENTIST	Receptionist	**0.53 (1.34)**	-1.35 (1.42)	-1.22 (1.52)	**1.74 (1.13)**
	" "+nurse	-0.50 (1.39)	**0.46 (1.78)**	N/A	1.64 (1.19)
	p-value	**0.0004***	**<.0001***	N/A	0.686

TABLE 4.3: M (SD) for each storyboard in the *Re-embodiment and Expertise* set (*C3-Re*). Values in bold with * are $<$ .05. The higher value of each measure for each storyboard pair is in bold text.

Air travel: different specializations in a single domain.

In one version of this storyboard, a boarding agent robot checked a user in for their flight, and then re-embodied to co-pilot the plane. In the other, the boarding agent robot re-embodied into the seat-back entertainment system while a different robot piloted the plane. For both versions, many participants welcomed the idea that re-embodiment could be comfortable if done right. Specifically, **participants commented that they would be comfortable with the situation *so long as* the robot was sufficiently capable of the skills required to do both jobs.** In the entertainment system version, some people talked about this dual competence as something they assumed the robot would have. In the pilot version, however, competence at both tasks was talked about as something that would have to be argued for or proven: P153 said, "If the robot is competent at both I don't see a reason why he shouldn't be able to do both jobs. The question comes in how competent he can be at piloting, especially in emergency scenarios." P72 said, "We tend to associate low-expertise jobs with a lack of competency in high-expertise jobs [...] could make some feel less confident." The notion of *if the robot can do multiple jobs, then it should* appeared in responses to the entertainment system version of this storyboard—and the other re-embodiment storyboards—as well, but these responses largely lacked qualifying comments that implied doubt about the ability to do multiple jobs.

Of the 17 participants who believed the agent embodying a boarding agent and a co-pilot should *not* take on multiple responsibilities (gave ratings to the right of zero), 10 called attention to the large difference between the two jobs. Participants explicitly called out **worries about risk and physical safety** as causes for concern: P22 said, "If they can't focus on their job and get mixed up, that could be disastrous," and P75 said, "More opportunity for something to go wrong. This especially applies to important responsibilities like piloting an airplane where there could be loss of life if something were to go wrong." The *improvement over the typical service experience* ratings were higher in the entertainment system version than in the pilot version. Participants also had more concern with the robot serving in both roles in the pilot scenario (see Table 4.3). Both scenarios introduced a pilot robot (see Figure 4.1), which suggests that the *re-embodiment* aspect was what raised concern and discomfort.

Dentist's office: empathy and training.

In the other three contexts for C3, one robot only ever took on a maximum of two roles in a single storyboard. In the dentist storyboards, the robot either took on two roles (parking assistant, then receptionist) or three roles (parking assistant, then receptionist, then dental nurse). Overall, the qualitative responses to this scenario looked similar to those from the air travel scenario. In general, people did not take issue with the same robot performing multiple roles. Those who did had concerns about expertise (e.g., P83 said a dental nurse was "a more specialized job") and risk of unexpected events (e.g., P38 said, "The job of dental nurse should be done by a human so that they can monitor pain or anomalies when cleaning.") As in the air travel scenario, a few participants expressed strong distaste without specific cause. The version in which the robot did *not* re-embody into the dental nurse was perceived as a significantly larger improvement over the typical service experience. People were also less concerned about the robot taking on multiple responsibilities when it did not serve in the role of the dental nurse (see Table 4.3).

Personal care: similar domains, but different specialties.

In one version of this storyboard, a hairdresser robot gave a user a haircut, and then re-embodied into a masseuse robot to give a massage. In the other version, it instead re-embodied into a physical therapist robot to consult with the user about their therapy regimen. Regardless of whether the robot became a physical therapist or a masseuse after first being a hair stylist, many people were comfortable with this re-embodiment. Participants said that **it made natural sense to them for one robot to do both of these tasks because they play similar roles in service users' lives**, and because interaction with people doing both of these tasks often looks similar. When people did not like the robot doing both roles, they did not call attention to either of the individual tasks as being specifically problematic. Rather, they noted that the tasks themselves were vastly different. This strikes a contrast with the air travel and dentist scenarios, in which robots re-embodying into a pilot and a dental nurse were seen as specifically off-putting.

Here, participants thought the service did a better job of creating a multitasking agent, believed the agent to be more competent, and were more comfortable with the robot serving in both roles in the masseuse scenario than in the physical therapy scenario (see Table 4.3). These findings suggest that people feel uncomfortable with re-embodiment when (1) one of the roles is high-risk, or (2) when the two roles are vastly different from each other and are taken from different service domains.

Overall, the findings related to C3 suggest that the appropriateness of re-embodiment accompanied by a change in expertise is determined in part by people's expectations about what tasks are typically done within the same domain, and in part by the perceived risk level of certain tasks. If a service robot is re-embodying within the same general domain, but will take on a new expertise, then the *kind* of new expertise impacts people's acceptance of the re-embodiment. If a re-embodiment would result in a new expertise that requires intense, specialized training and/or comes with a high perception of risk, then it is likely to make people uncomfortable. If the new expertise does not seem so specialized, then re-embodiment is likely to be perceived at least neutrally, if not positively.

4.4 Discussion

Our study revealed novel and critical insights about the way robots should and should not behave in service contexts. By situating three concepts from prior work in several different contexts, we were able to draw comparisons across different service settings. Our study was motivated by knowledge that context matters when robots identify people and re-embody. In this study, we gained specific knowledge of *how* context matters when robots identify people and re-embody. We organize our discussion around the three claims, and propose specific design recommendations for each (Table 4.4). In addition, we offer reflection on our use of an online storyboard study as one way of advancing knowledge from UE studies.

Design Topic	Service Setting	Finding	Design Implication
Identification by a robot	Personal appointment services	Being identified by a robot in a personal appointment setting can improve the service experience.	*If personalization is required for the service, a robot should identify the user. Otherwise, service robots should avoid identifying people.*
	Non-personal appointment services	Being identified by a robot in a non-personal appointment setting is perceived as creepy.	
	Personal and non-personal appointment services	Being identified by a robot that appears to use facial recognition is especially uncomfortable.	*Service robots that identify customers should not do so using facial recognition.*
Re-embodying robots	Roles that are perceived as managing high risk situations	Robots that re-embody in high-risk situations are perceived negatively and as unsafe.	*In low-risk service contexts, robots can re-embody to provide a better service experience.*
	Roles that are perceived as managing low-risk situations	People are accepting of robots that re-embody in low-risk service contexts.	
	Several tasks within one domain	Robots that re-embody for different tasks that are clearly in the same domain are perceived more positively.	*If robots fulfill several tasks in one larger domain, re-embodiment can improve the service experience. A robot should not re-embody to do tasks in different domains.*
	Several tasks in different domains	Robots that re-embody for tasks across different domains are perceived more negatively.	

TABLE 4.4: Implications for designing identifying and re-embodying robots for different service settings.

4.4.1 Claims

C1-Pers: *People will be bothered when a robot identifies them in a service where they would not expect to be identified.* There is variation on what sort of personal identification is and is not okay. Individual differences and cultural differences likely play into whether or not having personal information said aloud by a robot in a public place is acceptable or not. Additionally, the same individuals may welcome or oppose being identified in different contexts, and for different reasons. *Expectation to be identified* drives the *appropriateness* of service robots identifying people as part of their interaction design. Rather than a dichotomy, *expectation to be identified* is likely a spectrum. Gaining a full theoretical understanding of this spectrum would take additional research that is beyond the scope of our project, but our findings allow us to identify some possible important points along it. Additionally, people's comfort with being identified by a robot in a given context seemed closely tied to expectations of a human employee in that same context. However, something that differentiates robots from people is how the data is stored. Even when it is acceptable for service robots to have immediate access to a customer's complete profile, many people still object to automatic facial recognition. We therefore recommend that robots designed for a welcoming experience identify service users through opt-in, non-biological identifiers of customer history (e.g., a linked cell phone app, opted-in bluetooth, loyalty card).

C2-Pers: *People will be less bothered by being identified when a robot uses customer profile data to deliver something of value.* We expected that added value (in terms of some specific reward or benefit) would mediate the effects of creepiness when robots overtly identify people in public. In the superstore storyboard, the value was information; any positive effects of this were undermined by the extreme intrusiveness of user profiling. In the fast food and movie theater storyboards, the value was monetary. Here, the *added value* storyboards were perceived more positively than the *no added value* ones. We can speculate that when there is not a direct financial benefit involved, acceptance of a robot keeping and reciting personal information is likely more about *relevance* than value. However, because we did not compare any other types of value, we still do not know enough to make design recommendations about this claim.

C3-Re: *People will not want robots to re-embody when it involves a large change in social role or expertise.* The scenarios involving re-embodiment demonstrated concerns about multitasking, consistent with previous work (Luria et al., 2019). Participants in favor of re-embodiment noted the value of multitasking in providing familiarity and a seamless experience for the customer, and increased efficiency for the service. They recognized that robots could theoretically multitask better than humans, and believed that since they *can* take on multiple roles, they *should* take on multiple roles. However, they were concerned about multitasking robots when the jobs required very different skill sets.

A factor that emerged in our qualitative analysis, that we did not deliberately set out to manipulate or measure, was the amount of risk involved in each interaction. Two of our contexts were "high stakes", involving high levels of personal, physical risk, albeit of different kinds: inadequate service while on an airplane or at the dentist has the potential for disastrous results. One context involved some, but less, personal risk: a bad haircut, massage, or physical therapy appointment can have negative consequences, but these are usually not extreme or lasting. One did not involve any sort of physical interaction at all: while bad concierge information does pose some risk, this is usually a trivial concern compared to flying or medical anxieties. Responses to open field questions suggest that perceived high risk may be a primary driver of acceptance of agent re-embodiment into robots with different roles.

All in all, these findings suggest that perceived expertise matters for the acceptability of re-embodying service robots. Specifically, when robots perform different tasks in the same domain, perceptions about expertise will shape users' comfort with re-embodiment. Rather than a single scale, perceived expertise is likely a complex topographical space influenced by multiple constructs. For example, prestige may vary with expertise, or be a distinct concept from expertise. Likewise, risk may be tied to expertise when in dangerous settings due to perceived training and preparation.

4.4.2 Future directions

Our findings reveal directions for future work. In the C1 and C2 storyboards (about robots identifying people), we looked at a total of eleven specific contexts in which service robots might use personal information about customers in publicly observable interactions. We compared personal settings to impersonal ones, and situations where identification added value to situations where it did not. Across both of these comparisons, we concluded that people really do not want to be identified by robots for the sake of friendliness or customized recommendations; rather, they only want robots to identify customers when they *need* to confirm their identity to provide the service. There is also likely some gray area between "needing" to confirm someone's identity and not needing to (e.g., at a gym or an auto shop). Here, people may take a mostly neutral stance on the issue, and quickly brush aside or forget about slight discomfort or slight satisfaction with the behavior. Future work could seek to investigate this finding with a rigorous experimental approach in order to identify specific contexts where personalization and identification are widely desirable. This could factor into a taxonomy of contexts for personalized service robot interaction. The findings related to identification also suggested that there may be a difference between people's discomfort with robots storing their information and their discomfort with robots displaying their information where others can see or hear it. What kinds and what amount of information is okay for robots to *store* versus to *say*—and how this varies across services—remains an open question.

Through the C3 storyboards (on re-embodiment into robots with different roles and expertise), we began to explore possible definitions and impacts of expertise, role, status, and prestige for robots and AIs. Because artificial agents are not limited in the same ways as people are (e.g., they can exist in multiple places at once, they can have perfect memory, and they do not require as much time as humans do to "learn" how to demonstrate skill proficiency), these concepts will likely have different meanings for robots than they do for people. They undoubtedly will shape people's impressions, comfort, and trust differently when exhibited by robots. The human-robot interaction and service design communities could benefit from a deeper theoretical understanding of these social constructs as they apply to service robots, and we recommend that future work interrogate this. Finally, in the future, HRI researchers might consider exposing participants to in-person experiences derived from the scenarios in our storyboards via User Enactments and Wizard-of-Oz methods. A more personalized, higher-fidelity experience may reveal additional new insights on roles and expertise.

4.4.3 Reflection on methodology

In this study, we experimented with a novel method: speed dating with storyboards deployed to dozens of participants online. As with many design research methods like User Enactments and workshops, the method of speed dating with storyboards is most often used for in-person research, where a relatively small number of participants give detailed feedback on a few related design concepts. The work in this paper was done during the COVID-19 pandemic, when in-person interviews were not feasible. While the inability to conduct in-person research of course had its detriments, it also brought to light the promise of an innovative, mixed approach that combines aspects of multiple established methods. Prior work had provided us with preliminary knowledge about themes to touch on and comparisons to probe. We were able to use that knowledge to create targeted examples of personalized service robot interaction and re-embodiment, while still sampling a broad design space by way of the large number of storyboards (22 in total). By framing our study to participants as visionary rather than scientific and keeping the narratives simple and lightweight, we were able to encourage suspension of disbelief and open-ended reflection, which generated additional exploratory insights. At the same time, collecting both quantitative (i.e., Likert ratings) and qualitative (i.e., free response) data from a larger number of people than usually participate in speed dating studies allowed us to determine the prevalence of patterns and trends and examine the effect of specific service contexts. This narrower scope facilitated concrete design guidelines that would not be defensible if driven by an entirely open-ended design exploration involving just a few people. We recommend that other researchers consider adapting this method of online speed dating with storyboards for research that seeks intermediate-level knowledge on early, evolving design concepts.

4.4.4 Limitations

Several limitations of this work should be noted. The first comes from our sampling method—we only used a single recruitment platform, and we restricted participation to people in the U.S. and Canada. Therefore, the perspectives represented in our study are limited to those of a relatively small number of people, and may not reflect those of demographics not represented in our sample. Second, our study only used self-report measures. People may not be able to accurately predict their actual behavior or desires when judging imagined interactions with imagined robots. However, our method still provides insight into what they *value* in interactions with service robots. Finally, our stimuli were short vignettes that participants responded to in-the-moment. People's perceptions may shift over time with continued use. Determining the effects of long-term interaction, real-world interaction, willingness to use the service again, and individual and cultural differences (e.g., individualist vs. collectivist orientation, see Triandis, 2001) are additional promising directions for future research.

4.5 Summary and contributions

In this work, we built on findings from low-fidelity studies on behavior designs for service robots. Our goal was to deepen our knowledge through a more structured, mid-fidelity study: We tested several storyboards that each addressed a single claim in a particular service situation. Finally, this work contributes an example of how knowledge from initial exploratory research can be advanced.

The work described in this chapter makes the following contributions:

- Our findings inspire specific design implications for creating appropriate robot identification and re-embodiment behaviors based on the service setting (see Table 4.4).
- This work contributes an example of how two exploratory studies that assess similar design concepts in vastly different ways can complement each other. In the previous study (Reig et al., 2020, described in Chapter 3), we exposed a smaller

number of people to a richer set of immersive experiences, and collected detailed feedback. In this study, we collected data from a large number of participants based on a large number of low-fidelity stimuli deployed in a medium-scale online study. This allowed us to test specific questions and comparisons that arose in the first study.

Part III

Agent Identities and Collaboration Across Physical Embodiments and Task Domains

Chapter 5

Perceptions of Multi-Robot Failure Recovery Strategies

Much of this chapter was previously published as the scientific article:

Samantha Reig, et al. (March 2021.) "Flailing, Hailing, Prevailing: Perceptions of Multi-Robot Failure Recovery Strategies". In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*.

This chapter explores how an agent identity might mediate interactions between one individual and multiple robotic embodiments. When a human interacts with a robotic embodiment, the use of social cues to communicate states, needs, and processes is crucial. It is especially important during cases of failure, which can have lasting effects on perceived competence and trustworthiness (Desai et al., 2013; Desai et al., 2012; Morales et al., 2019). Sometimes, social cues will be intended to solicit help from human collaborators (e.g., Morales et al., 2019; Knepper et al., 2015; Tellex et al., 2014; Fong, 2001) or bystanders (e.g., Weiss et al., 2010). In these cases, they will be critically important to both robot function and human-robot relationships. In other cases, robots may recover autonomously without seeking human intervention, but they will still need to communicate to humans to repair trust and relationships (Kwon, Huang, and Dragan, 2018a).

When multiple robots work together, there may be cases in which a single robot experiences a failure from which it cannot recover sufficiently quickly (e.g., signal loss) or at all (e.g., severe hardware damage). One possibility is that the failure ends the task. However, it is also possible that the failed robot could find a way to resume the task (e.g., by downloading an update that improves its vision) or even hand the task off to another robot to complete. In these situations, will a violation of trust in the *robot system as a whole* be best repaired with a single robot that demonstrates resilience, or with a second robot that does not have the stain of a prior failure on its record? Could the software intelligence of the first robot re-embody another physical embodiment to achieve the best of both worlds?

To examine possible effects of recovery strategies with many participants, we designed an online study that showed videos of a package delivery scenario where robots carried boxes from point A to point B. This is similar to a paradigm from (Kim and

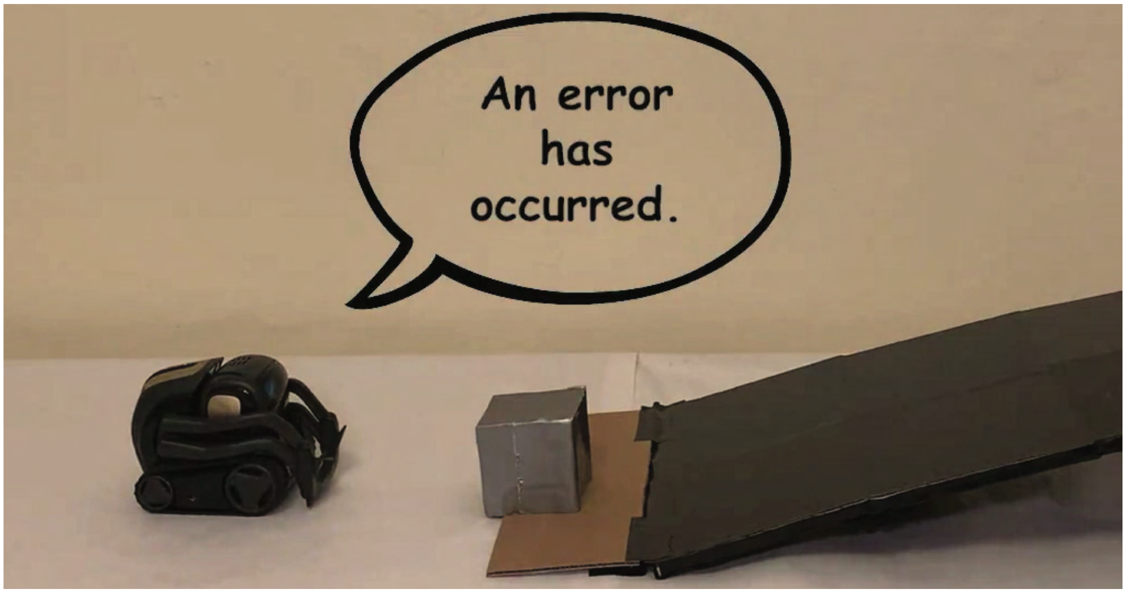

FIGURE 5.1: A robot drops a package at the bottom of a ramp.

Hinds, 2006) in which participants cooperated with a delivery robot in an assembly task.

5.1 Study A

5.1.1 Study design

This study had a between-subjects design with four conditions (one video per condition). Each video involved a small robot attempting to carry a small, solid, grey cube (a "package") from a starting point to an ending point. We used two Vector (*Meet Vector* n.d.) robots from Anki/Digital Dream Labs. These are small robots that have expressive, pixelated eyes and a bulldozer-like form. Each robot has a lift that is capable of picking up and placing down small objects. The robots used spoken natural language to explain what was happening. We also included speech bubbles to help participants understand the dialogue. We chose to use speech bubbles rather than captions because they could be placed next to the correct robot and thus be part of the scene.

All videos began the same way. First, the robot picked up the package and said, "Beginning package delivery." Then, it drove the package across a flat surface toward a ramp. At the bottom of the ramp, the robot swiveled back and forth, reversed, and put the package on the ground. It said, "Package dropped." After attempting to recover the package (by moving toward it and raising and lowering the lift), it reversed again, and declared: "Cannot recover package. Delivery failed. An error has occurred." Then, one of four recovery conditions was executed to complete delivery of the package.

- ***Update:*** *One intelligence, one robot. After a robot experienced a failure, it fixed the problem and then completed the task.* After acknowledging the error, the robot said, "Let me update my software," and drove back to the starting point. It then turned away from the camera and then back toward it, and it said, "The problem is fixed. I will not experience the same error again."

- ***Call:*** *Two intelligences, two robots. After a robot experienced a failure, it called a second robot that replaced the first one and completed the task.* After the first robot acknowledged the error, it said, "Let me call another robot," and drove back to the starting point. A second robot entered the frame, and said, "I will not experience the same error as the previous robot."

- ***Sense:*** *Two intelligences, two robots. After a robot experienced a failure, a second robot noticed the problem and replaced the first robot to complete the task.* After the first robot acknowledged the error, it drove back to the starting point. A second robot entered the frame and said, "I will take over from here. I will not experience the same error as the previous robot."

- ***Re-embody:*** *One intelligence, two robots. After a robot experienced a failure, it re-embodied (moved its intelligence to) a different physical robot to complete the task.* After acknowledging the error, the robot said, "Let me move my brain over to a better robot body," and drove back to the starting point. Its eyes and face went dark. A second robot entered the frame and said, "The problem is fixed. In this robot body, I will not experience the same error again."

At this point, the recovery robot (the same robot in the Update and Re-embody conditions; a second robot in the Call and Sense conditions) drove to the package, picked it up, and said, "Beginning package delivery." Then, it drove the package to the top of the ramp, placed it down, backed away from it, and said, "Delivery complete." All four conditions followed the exact same narrative up until the failure, and they resumed similar narratives after the recovery. Figure 5.1 shows an example of the failure event[1].

A pilot study with 154 participants on Amazon Mechanical Turk confirmed that (1) the package drop was perceived as a failure; (2) a successful robot was perceived as more trustworthy ($F(1, 152) = 37.76, p < .0001$, Cohen's $d = .97$) and competent ($F(1, 152) = 17.78, p < .0001, d = .68$) than a failing robot; and (3) they accurately understood the speech.

5.1.2 Hypotheses

We predicted that the recovery method used after a failure would impact participants trust. Prior work suggested that robots can recover from negative associations brought about by mistakes during sustained interactions using socially appropriate behaviors (Lee et al., 2010). Prior work also suggested that re-embodiment is perceived as a desirable and efficient design (Luria et al., 2019; Reig et al., 2020) and that identity migration positively impacts social perceptions (Tejwani et al., 2020). Thus, we predicted:

- **H1** Participants will have higher **trust** in a robot system following a **Re-embody recovery** than following an Update recovery.

- **H2** Participants will perceive a robot system that uses a **Re-embody recovery** as most **competent**.

[1]Full videos are included in the Supplementary Materials of the published paper.

Research on groups and teams of robots (e.g., Scheutz, DeLoach, and Adams, 2017; Gervits, Fong, and Scheutz, 2018; Gervits et al., 2020; Ma et al., 2018) informs our hypotheses regarding two-robot recoveries.

- **H3** Participants will have higher **trust** in a team of robots when the second robot **senses the first's failure** than when the first robot calls the second.
- **H4** Participants will perceive higher **competence** in a team of robots when the second robot **senses the first's failure** than when the first robot calls the second.

Our final hypothesis follows from the suggestions by previous work (Oistad et al., 2016) that favorable social perceptions of robots increase willingness to work with them in the future.

- **H5** Participants will report a greater **desire to use** the system in the future when they perceive it to be more **warm and likeable**.

5.1.3 Measures

Our assessments included a mix of questions from prior work and questions written for this study. The response format of the closed-ended questions was 5-point (*attitudes toward robots in general*), 7-point (*trust*), and 9-point (*competence*, *warmth*, *likeability*) scales.

Validation questions. To confirm that participants perceived the failures and recoveries as intended, we asked open-ended questions about their interpretations of the robots behavior during the task. We also included two attention checks that all passed.

Trust in the robot system. We evaluated trust through self-report measures. Participants were asked to answer several questions modified from the Jian scale (Jian, Bisantz, and Drury, 2000) and a few additional questions that we created specifically for this study.

Social attributions to the robot system. We used a subset of the 18-item Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017) to measure perceptions of *competence* and *warmth*. We analyzed both of these two factors and their individual items to examine more specific traits. To measure likeability, we used three Likert-type items inspired by words from the GODSPEED likeability subscale (Bartneck et al., 2009b).

Attitudes toward robots. We included five Likert-type items to obtain judgments of overall trust in robots, perceived helpfulness of robots, interest in robots, and perceived personal importance and societal importance of robots. Four of these were modified from a scale proposed (but not validated) in prior work (Reig et al., 2018). One, pertaining to overall trust, was new as of this work.

5.1.4 Procedure

Because some pilot responses from Amazon Mechanical Turk users suggested that people had glossed over some questions, we conducted the study on Prolific.co, which is a survey research platform with users who are used to longer-form studies. We described the task as gathering impressions of a prototype of a robotic package delivery system. Potential participants were redirected to Qualtrics for the study. After providing informed consent, participants were semi-randomly presented with one of the four videos (Update, Call, Sense, or Re-embody).[2] Below the video, participants were asked if the system experienced a failure and how it recovered from that failure. They then answered the questions about trust (presented in a random order), social attributes (in a random order), and attitudes toward robots. Then, they answered demographic questions, including about their age, gender, languages, employment, experience with computers and robots, and an open-ended question meant to capture additional demographic information. Finally, participants had the option to provide feedback about the study.

5.1.5 Participants

A total of 403 people participated in this study. There were 100 participants in the Update condition, 100 in Re-embody, 101 in Call, and 102 in Sense. To be eligible for the study, Prolific users had to be 18 years of age or older, be located in the U.S. or Canada, be proficient in English, and have a previous submission approval rate of at least 95%. Participants ranged in age from 18 to 78 years ($M = 31.25, SD = 10.89$). 162 were female, 234 male, 5 were other genders, and 1 did not specify a gender. They had a variety of professional backgrounds, including engineering, medicine, psychology, art, and sales. They generally had some experience using computers and little experience using AI personal assistants and robots (on a 7-point scale with 7 being more use, computers: $M = 6.70, SD = 0.70$; AI assistants: $M = 2.88, SD = 1.98$; robots: $M = 1.98, SD = 1.45$). 251 owned a pet, 257 owned an AI assistant, and 57 owned a robot. Participants took an average of 14 minutes to complete the study (min: 5, max: 45, median: 12) and were paid 2.50 USD each. Our study was approved by an Institutional Review Board.

5.2 Study A Results

Explanations of the failure and recovery accurately reflected the differences between the robot behavior in the different conditions, suggesting that the conditions were interpreted as intended. We analyzed the data using a linear model fit with REML.

The *trust* questions were correlated at Cronbach's $\alpha = .89$. The RoSAS *competence* items had $\alpha = .88$, and the *warmth* items had $\alpha = .90$. We treated these as factors. We analyzed *likeability* as an individual item because *meanness* and *friendliness* only weakly correlated with it. The *attitudes toward robots* questions correlated strongly ($\alpha = .85$) and were treated as a factor.

We included the *attitudes toward robots* questions to understand whether preexisting associations or biases had an effect on our dependent variables. In an exploratory analysis, we found that the factor had a significant effect on *trust*, *warmth*, perceived *competence*, and *likability*, $p < .0001$ for all variables. We placed these items at the end of our study rather than at the beginning in order to prevent priming the participants to rate the videos according to the immediate availability of their preexisting attitudes rather than our manipulation. We were concerned that the attitude questions could have been affected by our manipulation, thus invalidating attitude as an independent

[2]The video only allowed for pause and play; participants could watch the video more than once, but could not fast forward, rewind, or change the playback speed. Participants were told that they would only be able to watch the video straight through and that they could not proceed to the next questions until an amount of time equal to the video duration elapsed.

variable. We ran a nonparametric Wilcoxon rank sum/Kruskal-Wallis test to check for this. We did not find any significant effects of condition on attitudes (in fact, all means were $M = 3.7$). After confirming that it was not affected by condition, we included attitude in our model as a covariate. We used Tukey's Honest Significant Difference (HSD) test for post-hoc comparisons.

5.2.1 Trust in the robot system

We found a main effect of Recovery method on *trust*, $F(3, 395) = 3.16, p = .025$. Post-hoc pairwise tests revealed that trust was higher in the Update condition ($M = 4.03, SE = 0.10$) than in the Sense condition ($M = 3.67, SE = .10$). Because there are different dimensions of trust, we also looked at the individual items from the scale. We found a main effect of Recovery condition on perceptions that the system was *reliable*, $F(3, 395) = 2.71, p = .0345$. Post-hoc tests showed that the Re-embody recovery ($M = 4.19, SE = .14$) was rated higher than the Sense recovery ($M = 3.67, SE = .13$). We also found a main effect of Recovery condition on desire to use the system in the future, $F(3, 395) = 2.99, p = .031$, which was higher for Update ($M = 4.39, SE = .15$) than Sense ($M = 3.83, SE = 1.69$). We did not find trust differences between Update and Re-embody, so **H1 was not supported**. We also did not find any trust differences between the Call and Sense conditions, so **H3 was not supported**.

5.2.2 Perceived competence of the robot system

For perceived *competence*, we found a main effect of Recovery method, $F(3, 395) = 3.25, p = .022$. In particular, Update ($M = 5.81, SE = .14$) was perceived as more competent than Sense ($M = 5.22, SE = .14$). We also found an interaction effect of Recovery method and attitudes toward robots, $F(3, 395) = 3.31, p = .020$. Higher scores on the attitudes index combined with a Re-embody recovery led to higher perceptions of competence, $p = .046$. This did not directly support H2, but it did suggest that re-embodiment was perceived as a more competent design by participants who had positive attitudes toward robots. We analyzed the individual items for the competence scale as well, and we found a main effect of Recovery condition on perceptions of the system as *knowledgeable*, $F(3, 395) = 3.56, p = .015$. Specifically, Re-embody ($M = 5.81, SE = .20$) was perceived as more knowledgeable than Sense ($M = 4.97, SE = .20$). Re-embody was higher than Sense, but not Call, and only on one item of the *competence* construct; this meant that **H2 was partially supported**. We did not find differences for competence between Call and Sense, so **H4 was not supported**.

5.2.3 Social attributions to the robot system

We did not find any effects of our manipulation on *warmth* or *likeability*. However, we found an interaction effect of Recovery method and attitudes toward robots on likeability, $F(3, 395) = 3.94, p = .009$. Higher attitudes scores combined with a Re-embody recovery led to higher likeability, $p = .023$. Desire to use the robot system in the future was moderately correlated with perceived *warmth*, $r = .37$ and with *likeability*, $r = .45$, both $p < .0001$, **supporting H5.**

5.3 Study A discussion

In Study A, we predicted that a Re-embody recovery would result in the highest perceived trust and competence, and that Sense would be perceived as more trustworthy and competent than Call. Three of our hypotheses were not supported, and one received only partial support. In general, Re-embody was not an improvement over Update, and Sense was not an improvement over Call. Instead, the common thread across our findings was that Update was perceived most favorably, and particularly more favorably than Sense.

To explore possible explanations, we looked at the qualitative data, which consisted of reflections on the recovery, explanations of the trust and social attribute ratings, and general feedback. We noticed that participants anthropomorphized the robots (e.g., "He wants to update his software so he won't experience the same error again,"–P391) and viewed them as cute (e.g., "The voice was very cute and so were its little eyes,"–P51). However, they were not willing to associate robots with words meant to measure perceived *warmth* because "robots do not have emotions" (many participants). In particular, when participants saw two robots, they especially anthropomorphized the first robot and thought it "made you feel bad for the little guy when he failed" (P210). This endearing failure caused them to see the first robot more positively when it recovered. For example, P121 said, "It didn't get grumpy while experiencing an error but instead acted promptly and made an immediate effort to find a solution." P270 said, "I honestly thought the first robot looked very distressed [...] The little fella looked cute as hell and I was touched." In contrast, participants viewed the second robot negatively when it took over. P288 said, "I felt sad for the first robot." P258 said, "The second robot was 'mean' by dismissing the first robot, and I was weirdly almost rooting for it to fail."

We reason that participants anthropomorphized the first robot and then favored Update because it was the condition in which the first robot showed the most agency: it failed, was able to repair the error on its own, and then continued the task successfully. Conversely, in the Sense condition, the first robot had the least agency: it simply stopped and waited for another robot to come and take over. Besides forming an attachment to the first robot, participants also felt that the need for a second robot made the system as a whole less reliable. For example, P233 said, "Ideally, there should be no need to depend on a second robot," and P235 said, "The first robot should have made another attempt."

We also noticed a pattern where participants commented that they based their ratings of trust entirely on the fact that the first robot failed to deliver the package on the first try. For example, P7 said, "It looks like it's in early testing, and it doesn't seem too reliable as the first one failed the simple task." The timing of a trust violation influences changes in trust (Desai et al., 2013; Desai et al., 2012). In our study, there was no "burn-in period" for building up trust before the error occurred. It is possible that the effects of our manipulation were dwarfed by the effect of seeing only a single, failed first attempt at delivery.

Results may have also been impacted by participants taking the perspective of the package *recipient*, rather than that of someone who *worked with* the robots. Many participants mentioned that they would not be willing to trust the system enough to use it until it showed major technical improvement (e.g., "I'm not confident that it could be trusted in more complex, real-world settings,"–P317; "I would likely not use [it] in case of future errors that could not be automatically resolved,"–P365). Several participants mentioned concerns that the robot(s) would not be able to handle stairs (e.g., P60, P87, P140) or bad weather (e.g., P53, P209), or that packages would be subject to theft (e.g., P61, P91, P351). From the vantage point of an end-user who would only ever *see* such a system if it succeeded, people were hesitant to view it as trustworthy and competent if it could not successfully perform its task even once.

This study provided evidence that participants did not make social attributions to the robots despite anthropomorphizing them, that people generally preferred a one-robot recovery over a two-robot recovery, and that participants formed impressions of the robot(s) from the perspective of an end-user or customer rather than a collaborator. With these new insights, we conducted another study to better understand these findings.

5.4 Study B method

We adapted the method from Study A. We used the same videos, recruitment platform (Prolific), and survey template (in Qualtrics).

5.4.1 Methodological adjustments

In this section, we describe the changes from Study A. Methods not described here (e.g., recruitment, consent) remained the same.

Scenario framing. We revised the introductory blurb for the study to invoke a collaboration with the robots rather than receiving a service. It read: *"In this study, you will learn about and watch videos of a prototype for a robotic package delivery system. Imagine that you work with the robots that are part of this system. You are responsible for managing them as they coordinate to deliver packages. Because of various obstacles in the environment, they sometimes fail, but they have protocols in place to resume the task after a failure."*

Within-subjects design. To further examine differences in perceptions and attributions between "one-intelligence" (Update and Re-embody) and "two-intelligence" (Call and Sense) conditions, we used a within-subjects design. Each participant viewed all four conditions in a random order.[3] This also enabled us to ask participants to rank the four designs in order of preference.

Timing of the failure. We added a Baseline video in which a single robot successfully delivered the package on the first try. Thus, success was shown as a possibility and the first failure was not experienced as early. We expected this addition, along with the within-subjects design, to recalibrate participants' ratings of the system's trustworthiness and competence after recoveries.

Measures. The Study A findings about non-social treatment of the system as a whole, anthropomorphism of the first robot, and attributions of failure informed our measures for Study B.

Trust questions. We used the Muir trust scale (Muir, 1989) rather than the Jian trust scale (Jian, Bisantz, and Drury, 2000). The wording of the questions in the Muir trust scale is less evocative of relational aspects of trust, which makes more sense for a study

[3] Because the order was randomly chosen each time by our survey software, the 24 $\binom{4}{4}$ ordering conditions were not balanced. However, the number of times each Recovery condition occurred in each position was sufficiently distributed.

in which participants are not interacting with robots or viewing them socially. Prior work on failures in HRI has shown that both scales elicit similar ratings of trust (Desai et al., 2013).

Attribution of failure. We added a question about whether participants attributed the robot's failure to get up the ramp to a hardware problem, a software problem, both, or another problem. We asked this question for each condition.

Agency of the first robot. In Study A, the RoSAS *warmth* subscale was subject to a floor effect: participants did not attribute the descriptions of words like "emotional" and "organic" to the robots they saw in the video. However, they did anthropomorphize the first robot in their qualitative descriptions, and this seemed to influence their perceptions of the two-robot conditions. Therefore, we replaced the RoSAS *warmth* subscale with measures of *agency* and *anthropomorphism*. We used analogical statements from Ezer's robot anthropomorphism instrument (Ezer, 2008), items from Kozak et al.'s Mind Attribution Scale for perceptions of agency (Kozak, Marsh, and Wegner, 2006), and one new item ("The robot is capable of complex thought"). These instruments have been used in prior HRI work on robots in groups (Fraune et al., 2020).

5.4.2 Hypotheses

We approached Study B with a novel set of hypotheses. Because the Study A results implied that perceptions of the whole system were primarily shaped by perceptions of the *first robot*, we predicted:

- **H6a** Participants will perceive a robot that experiences a failure to have more **agency** when it **recovers on its own** than when it requires help from another robot.
- **H6b** Participants will perceive a robot that experiences a failure to be more **competent** when it **recovers on its own** than when it requires help from another robot.
- **H6c** Participants will have higher **trust** in a robot system in which one robot recovers on its own than in a robot system that uses a two-robot recovery.
- **H7a** Participants will have a greater **desire to work with** a system in which they perceive a failing robot to have more **agency**.
- **H7b** Participants will **prefer** a robot system that recovers using the **same hardware and the same software**.

We also tested the suggestion from Study A that participants formed an attachment to and "rooted for" the first robot's AI:

- **H8** A failure that is recovered with a **re-embodiment** will be perceived as a **hardware problem** (rather than a software problem) more often than will a failure that is recovered by the same robot without a re-embodiment or by a second robot.

5.4.3 Participants

We recruited 130 participants for this study, none of whom participated in Study A. Participants ranged in age from 18 to 64 ($M = 29.81, SD = 9.67$). 51 identified as

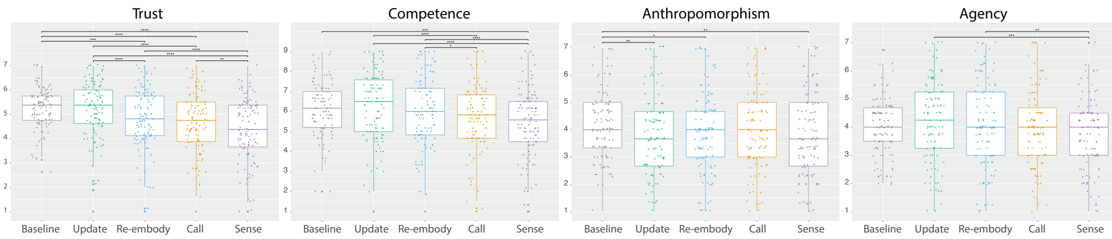

FIGURE 5.2: Box plots showing trust, competence, agency, and anthropomorphism for the Baseline video and each of the four Recovery conditions (Update, Re-embody, Call, Sense). Brackets marked * are significant at the .05 level, ** shows significance at the .01 level, *** shows significance at the .001 level, and **** shows significance at the .0001 level.

female, 57 as male, 1 as nonbinary, and 1 as agender. As in the first study, many different personal and professional backgrounds were represented (e.g., engineering, law, science, retail), experience with computers was high ($M = 6.75, SD = 0.65$), and experience with AI personal assistants and robots was relatively low (AI personal assistants: $M = 2.52, SD = 1.81$; robots: $M = 1.76, SD = 1.08$). 59 owned a pet, 73 owned an AI personal assistant, and 15 owned a robot. Participants took an average of 38.2 minutes to complete the study (excluding one outlier) and were paid 5.00 USD each.

We excluded data from 20 participants who (a) failed the attention checks, (b) perceived the Baseline video to have a failure, (c) did not perceive one of the failures to be a failure (this would have interfered with the way their impressions changed across conditions), or (d) used a mobile device (we could not prevent scrubbing the video for mobile viewing). This left us with a total of 110 participants.

5.5 Study B Results

The residuals were non-normally distributed, so we used Friedman tests and Wilcoxon signed-rank tests with a Bonferroni correction for post-hoc comparisons unless otherwise noted. Where possible, we report effect sizes with Kendall's W for Friedman tests and with r for post-hoc tests. We report sample medians as M.

The Muir *trust* scale had a Cronbach's $\alpha = .94$. The RoSAS *competence* items had $\alpha = .89$. We created a factor out of the analogical statements for *anthropomorphism*, which had $\alpha = .77$. Four of the five *agency* items had $\alpha = .77$. One of them, "The robot is capable of doing things on purpose", was only weakly correlated with the other items, so we excluded it from the agency factor.

5.5.1 Trust in the robot system

We found a main effect of Recovery method on *trust*, $\chi^2(4) = 98.8, p < .0001, W = .22$. Trust was significantly higher in Update ($M = 5.38$) than in Re-embody ($M = 4.81$), Call ($M = 4.75$), and Sense ($M = 4.38$), all $p < .0001, r > .48$. Trust was significantly higher in Re-embody than in Sense, $p < .0001, r = .45$, but there was no significant difference between Re-embody and Call. Also, trust for Call was significantly higher than for Sense, $p = .002, r = .35$. Finally, trust was lower in Call and Sense than in the Baseline ($M = 5.38$), $p < .0001$ ($r = .47$ and .65, respectively), and lower in Re-embody than in the Baseline, $p = .0007, r = .37$. Trust in the Update condition was not significantly different from Baseline. These results **support H6c**.

5.5.2 Perceived competence of the robot system

There was a small but significant main effect of Recovery method on perceived *competence*, $\chi^2(4) = 44.3, p < .0001, W = .10$. Specifically, perceived competence was significantly higher for Re-embody ($M = 6.00$) than for Call ($M = 5.83$), $p = .022, r = .29$, and for Sense ($M = 5.58$), $p < .0001, r = .46$, **supporting H6b**. Update ($M = 6.50$) had the highest rating and was also perceived as more competent than both Call and Sense, $p < .0001$, ($r = .44$ and .53, respectively), **supporting H6b**. There was no significant difference between Update and Re-embody, nor between Call and Sense.

5.5.3 Social attributions to the robot system

There was a small effect of Recovery on perceptions of the first robot's *agency*, $\chi^2(4) = 17.4, p = .0016, W = .04$. The robot was perceived to have more agency in Re-embody ($M = 4.00$) than in Sense ($M = 4.00$), $p = .002, r = .35$, and more agency in Update ($M = 4.20$) than in Sense, $p = .0002, r = .41$. There was also a small effect of Recovery on the *anthropomorphism* of the first robot, $\chi^2(4) = 22.90, p = .0001, W = .05$. The robot in Baseline ($M = 4.00$) was perceived as more anthropomorphic than the first robot in Re-embody ($M = 4.00$), Sense ($M = 3.67$), and Update ($M = 3.67$) ($r = .29, .36, .35$) but there were no significant differences between Baseline and Call ($M = 4.00$) or among the failure conditions. As such, **H6a was partially supported**. Desire to work with the system in the future moderately correlated with increased ratings of the first robot's anthropomorphism, Pearson's $r = .46, p < .0001$, and its agency, $r = .37, p < .0001$, **supporting H7a**.

5.5.4 Attributions of failure

We used Cochran's Q test to examine effects of Recovery condition on attributions of the failure, treating each possible attribution as a binary variable (1 if it was the participant's answer, 0 if it was not). There was a significant effect of Recovery on ratings of the failure as a hardware problem, $\chi^2(3) = 129.0, \eta^2 = .39$, as a software problem, $\chi^2(3) = 178.0, \eta^2 = .54$, as both, $\chi^2(3) = 60.9, \eta^2 = .18$, and as other, $\chi^2(3) = 24.8, \eta^2 = .08$, all $p < .0001$. We used pairwise McNemar tests for post-hoc comparisons. The failure was attributed to a hardware problem significantly more in the Re-embody condition ($n = 66$) than in the Update condition ($n = 1$), $p < .0001$. We also found that the failure was attributed to a hardware problem significantly more in Re-embody than in Call ($n = 23$), $p < .0001$ and Sense ($n = 13$), $p < .001$. These results **supported H8**.

5.5.5 Preference

A majority of participants ($n = 73$) ranked Update as their most-preferred recovery (Figure 5.3), followed by Re-embody ($n = 19$), Call ($n = 14$), and Sense ($n = 2$). Most participants ($n = 48$) ranked Sense as their last choice. Interestingly, Re-embody was also frequently the least-preferred recovery ($n = 37$). **H7b was supported.**

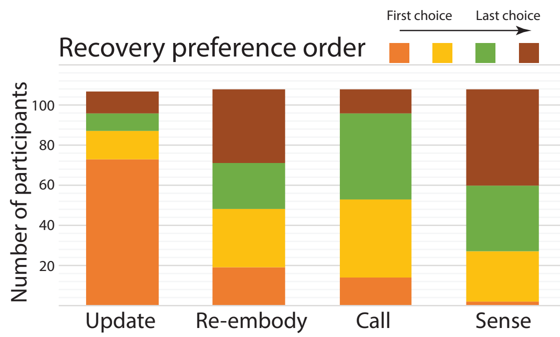

FIGURE 5.3: Most Study B participants rated Update as their first choice. Re-embody and Sense were commonly ranked last.

5.5.6 Other findings

We conducted additional exploratory analyses to look for effects of Recovery condition on the individual analogical statement items from (Ezer, 2008). We used the Skillings-Mack test to look for effects on on perceptions that the first robot was *like a pet* and *like a teammate* because some values were missing. We used Friedman's test for *like an assistant*. There was a main effect of Recovery on perceptions of the first robot as a pet, $\chi^2 = 9.65, p = .047$. Post-hoc Wilcoxon signed-rank tests revealed that the first robot was perceived as more like a pet in the Baseline than in Call, $p = .041, r = .30$, Re-embody, $p = .006, r = .32$, and Update, $p = .015, r = .29$, but not in Sense. There was also a small effect of Recovery on perceptions of the first robot as an assistant, $\chi^2 = 43.7, p < .0001, W = .10$. Ratings were higher for the Baseline than for all four failure conditions, all $p < .001, .39 < r < .50$. There was no effect of Recovery condition on perceptions that the first robot was like a teammate.

5.6 General discussion

The recovery strategies we tested compared a single-robot-single-AI recovery (Update condition), a multi-robot-single-AI recovery (Re-embody), and two forms of multi-robot-multi-AI recoveries (Call and Sense). We approached these two studies expecting to see a pattern in which the recoveries with more-efficient designs would be perceived more favorably. Instead, we found that people "rooted for" a robot that had failed: they perceived the system to be more trustworthy and competent in the single-AI Update and Re-embody conditions than in the Call and Sense conditions.

It is interesting that attachment to a single robot and perceptions of agency played a role in shaping trust and perceived competence despite relatively low ratings of warmth (Study A) and anthropomorphism (Study B). This suggests that people viewed the robots through a social lens despite claiming to consider them functionally. The field of HRI has long known that humans can form and benefit from bonds with machines despite *knowing* that they are machines that do not themselves have feeling. Nass' famous Computers Are Social Actors theory emphasized that social treatment of machines impacts human-machine relationships and occurs *independently* of true mind attribution and even anthropomorphism (Nass, Steuer, and Tauber, 1994). It follows that when robots experience damage or fail, their human partners will emotionally invest in their recovery. In fact, this has been reported in stories of soldiers whose life-saving robots have been damaged (Singer, 2009). Our results demonstrate a type of preference or attachment for the first robot to attempt recovery even in a non-interactive scenario. This raises an interesting question about how to rebuild trust in robots after failure and the relationships among failure recovery and form, agency, and anthropomorphism.

Taken together, our findings suggest that the software update recovery was perceived most positively overall. However, the re-embodiment condition—in which the same interactive AI continued the task by moving into a different physical robot—was a fairly close second on many outcomes (see Figure 5.2). This has implications for specialized, goal-oriented, and high-risk environments: Robots that work closely with humans in task-oriented settings might be designed to take on a social "software identity" that can persist across embodiments to maintain trust after unexpected errors and failures. Relatedly, in Study A, individual differences influenced how positively participants responded to the re-embodiment recovery. It is likely that the impact of re-embodiment recoveries on trust repair and human-robot relationships varies according to other individual differences as well. Socially interactive robots can be designed to behave differently when recovering after a failure depending on task domain, team dynamics, and personal traits of the current user(s). This is an opportunity area for future research.

5.6.1 Limitations

Our study was conducted on one recruitment platform with a relatively small sample from the U.S. and Canada. The perspectives in our results may be limited by the sample's demographics, and our findings may not generalize to other populations. All of our findings were based on self-report measures, which do not always correspond to behavioral metrics meant to assess similar variables (e.g., objective and subjective trust measures do not always correlate).

Additionally, it is possible that aspects of our video stimuli not related to the manipulation impacted the results. Making videos that varied only by the minimum amount of dialogue and robot movement necessary to differentiate the recovery strategies was an intentional choice to minimize possible confounds. However, it is possible that the videos were too alike, especially in the Call and Sense conditions, for participants to find them noticeably different. The use of the word "software" in the Update condition and "brain" in the Re-embody condition may have impacted perceptions of anthropomorphism, and results more generally. We intended for the Sense condition to be interpreted as one robot proactively helping another after detecting its failure, but participants may have instead interpreted this as the first robot implicitly summoning the second. A stronger signal of a proactive response by the second robot might have drawn a starker contrast between Call and Sense, which were perceived overall similarly in both of our studies.

We also used robots that were small and toy-like, and which many participants called "cute". Although the robots had a functional form, their expressive eyes, high-pitched voices, and use of natural language likely raised expectations about anthropomorphism. The study results might have been markedly different had we used a different robot. Even with the Vector robots, we might have seen different patterns if the robots' eyes had been hidden, or if the state had been conveyed through different signals (e.g., as simple messages on a scrolling text log).

Finally, our study is limited in that it sought insight into human-multirobot interaction but did not involve in-person human interaction with real robots. We found that supplementing our closed-ended survey questions with open-ended ones was particularly useful given this setup. Analyzing short-answer explanations of closed-ended questions facilitated the discovery of qualitative insights that might have emerged through interviews or observations in an in-person, laboratory setting. These insights helped us develop Study B, which was instrumental to the conclusions we drew from this research. Still, future work is needed to examine how people react and respond to multi-robot failures and recoveries during real-life interactions.

5.7 Summary and contributions

A robot's immediate response to a failure can have critical and lasting effects on trust and other HRI outcomes. Multi-robot systems have a number of options for how to recover from failures in ways that repair trust and other aspects of human-robot relationships, some of which involve using the same agent identity to re-embody into new robots. This study examined the effects of four failure recovery strategies on trust, perceived competence, and social perceptions of a multi-robot system. In an online study, participants watched videos of a robot that recovered from a failure by updating its software, by re-embodying into another robot, by calling for a second robot, or by getting assistance from a second robot that detected the problem. The findings have implications for human-robot interaction design during instances of failure as well as for human-multirobot-interactions more broadly.

The work described in this chapter makes the following contributions:

- We found that trust and perceived competence of a multi-robot system were highest when a single robot with a single identity recovered on its own.
- We found that a single agent identity re-embodying into a new robot brought about higher perceptions of trust and competence following a failure than a second robot with a separate identity.
- We found that observers attribute failures that are recovered using re-embodiment to hardware problem more than they attribute failures that are recovered using a second robot (with a second agent identity) to a hardware problem.
- Our study suggests that after seeing a robot system experience a failure, people will be more likely to want to work with it again if they perceive it to have more agency.

Chapter 6

Agent Affiliation, Reference Cues, and Roles in Smart Environments

Portions of this chapter were previously published as the scientific article:

Samantha Reig, et al. (June 2022.) "Theory and Design Considerations for the User Experience of Smart Environments". In *IEEE Transactions on Human-Machine Systems, Vol. 52, No. 3*.

In an increasing number of settings, multiple automated processes work together with various interfaces in what has been called a "smart environment". This situated and interconnected (and possibly interoperable) set of technologies—including robots, agents, and interfaces—provides useful services to the human inhabitants and users of the space. In homes, family members interact with voice assistants and smart devices that provide support for the activities of daily life. In factories and industrial settings, robots perform assembly, inspection, and other tasks, and they coordinate for optimal workflow with the support of sensors monitoring human activity. Three Astrobee robots now fly about the International Space Station to support astronauts with everyday tasks (and relieve them of some), and NASA is preparing for the presence of more intelligent systems aboard crewed and uncrewed spacecraft. In the future, these smart environments will become even more common, as well as more varied in terms of their goals, their tasks, their physical and interaction designs, and the ways in which they facilitate the interfacing of artificial and human intelligence.

With a shift in mentality and design from independent systems to connected or interdependent systems comes an increased need to study individual and group human-agent interactions. Here, too, agent identities may mediate several kinds of relationships that exist in these complex environments. For example, a single identity may be ascribed to all robots, cameras, speakers, screens, and other output devices of an environment, mediating the interactions between individuals and all embodiments. This chapter focuses on a simplified version of this paradigm. We conducted a study to investigate how an agent identity's mediation of an *embodiment-embodiment-individual* relationship may differ based on how it is embodied, with whom or what it is affiliated, and its expertise (which was a prominent theme in the work described in Part II). As this work focuses on multi-embodiment-multi-person interaction and draws heavily on the concept of smart environments, this chapter begins by providing some (re)framing of just what a "smart environment" is or could be.

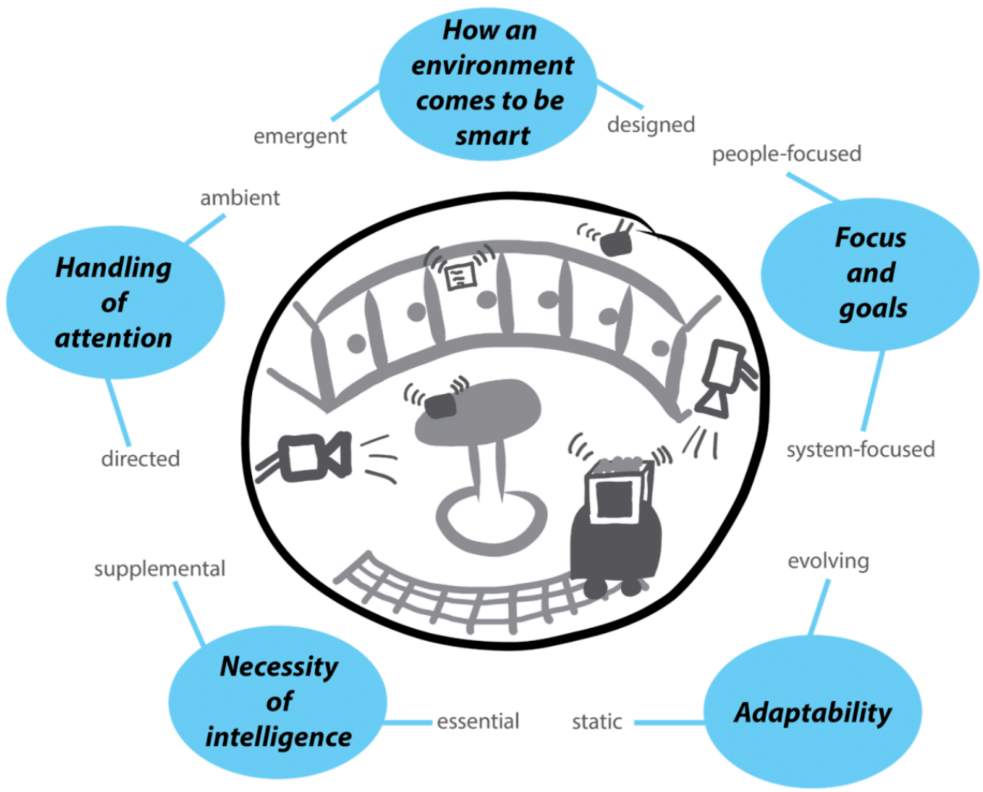

FIGURE 6.1: Five lenses for a systemic view of smart environments, with example variations on the lenses.

6.1 Broadening the lens and peering through new lenses: A systemic view of interactions in smart environments

Existing literature that addresses the user experience (UX) of smart environments (SE) consists mostly of works in four categories. Two of these are explicitly focused on smart environments. The first is conceptual visions: papers in this category articulate definitions (e.g., Das and Cook, 2006), grand challenges (Streitz et al., 2019; Stankovic, 2014) and design priorities (e.g., Gomez et al., 2019) for smart environments. The second is interface design: a number of researchers (e.g., Luria, Hoffman, and Zuckerman, 2017; Segura et al., 2012; Schiffhauer et al., 2016) have compared possible interfaces for interaction with an intelligent space. The other two informative areas of literature are not themselves focused on smart environments, but inform a UX understanding of them. The third category is work on social and interpersonal dynamics when multiple humans and/or multiple technologies interact in a group (e.g., Chaves and Gerosa, 2018; Fraune, Šabanović, and Smith, 2017). The fourth is user-appropriate system autonomy that facilitates accurately calibrated trust (e.g., Fallon et al., 2010) and prioritizes human autonomy (e.g., Jaschinski, 2014).

6.1.1 Lenses to a systemic view of smart environments

Across these bodies of literature, there is a theme of "people-empowering smartness" (Streitz et al., 2005; Streitz, 2019) and a sense that technology in smart environments must give inhabitants final say. This perspective is a necessary and critical (though not sufficient) North Star for design and ethics. However, it is not always specific enough to address in detail the many parameters under consideration for the architecture of a smart environment and its socially interactive components. A systemic view of smart environments—which applies multiple perspectives to their study, taking into account their reason for inception, their hardware, the modes and means of interaction they afford, their stakeholders, and their configurability—is needed to frame our understanding of the user experiences they will provide.

From the insights that we gleaned from the bodies of literature reviewed for this thesis, we derived five **lenses for the study of smart environments** that should be considered for research and development (see Figure 6.1). The process of identifying the lenses consisted of maintaining an annotated bibliography as we conducted our literature review and taking thorough notes of patterns in findings, underexplored relationships between bodies of literature, and perspectives that were missing around the concept of "a systemic view of the UX of smart environments". We reviewed and discussed these notes, reorganizing and synthesizing until we settled on the five concepts that we present in the following section. For example, our realization that most work concerned with HAI was situated in smart homes inspired us to differentiate "people-focus" from "system-focus" in environments. Our review of papers on user values and ethics highlighted the need for the *adaptability* lens, which focuses on user control and customization. What we cover in this section is not exhaustive; other perspectives on smart environment UX will likely be identified as technology and research evolve. Instead, the five lenses provide a foundation for discourse and research that position the **smart environment as the unit of analysis** and consider stakeholder values, fitting interface design, multiple users, and trust and autonomy of inhabitants.

How an environment comes to be smart

Do a conglomeration of smart speakers, vacuum robots, robot arms, and smart TVs that all operate within one room turn the room into a smart environment? Do a dozen different Alexa-enabled devices in a single room turn that room into a smart environment? Is a room that lends itself equally well to manual operation as it does to autonomous operation smart enough to be "smart"? The answer to all of these questions can be "yes", though the spaces they characterize are very different. It is useful to draw an explicit distinction between these types of environments.

Emergent. An emergent environment results from the accumulation (over time or at once) of a number of different "smart parts". It is likely to be created *by* or *with* its users rather than *for* them via the gradual and ad-hoc addition of new devices, agents, protocols, and other technologies. It also may be created for them by a third party that gradually develops the environment over time. Because it is not necessarily created intentionally, it lacks scheduled upgrades and intentionally-imposed constraints; therefore, it may evolve unpredictably. The mental model of an emergent environment is likely to be a collection of individual things that work together (even beautifully), but not a unified whole. A fully emergent environment would probably not be conceptualized as a single agent.

Designed. In a designed environment, smart parts are intentionally put together in a top-down design process. Because careful consideration is likely during the initial construction of this type of environment, the paths it might take to evolve over time are fairly predetermined. Rather than a collection of complementary parts, the mental model of a designed environment is likely to be more of a gestalt. A fully designed environment could be conceptualized as an agent, in and of itself, that controls multiple pieces of hardware and software.

Focus and goals

This lens to smart environment UX design centers on the numerous stakeholders of smart environments and their sometimes-conflicting values. In keeping with our theme of contrasting possible ways in which each lens can manifest in an individual smart environment, one way to conceptualize this possible tension and how it can be effectively accommodated is to differentiate between environments in which the driving goal is a "good" experience for the residents (for whatever is the immediate meaning of "good") and those in which the driving goal is something else. "Something else" could be anything that is not *primarily* or *solely* concerned with a good experience for the people inside—perhaps instead prioritizing a technical objective set by a third party, a service for a remote stakeholder, or the survival and success of the environment itself. We call the former (environments that exist for the comfort and happiness of inhabitants, as in smart home model) "people-focused" and the latter (environments that exist for something other than the inhabitants' experience) "system-focused".

Depending on a specific stakeholder's perspective, a smart environment or a subset of its components may appear system-focused or people-focused. For example, in a smart spacecraft, a non-habitation module that is mostly closed off to astronauts would appear system-focused to the astronauts (there to do a job at the command of mission control). However, from the point of view of the engineers on the ground who oversee that module, it would appear people-focused (there to serve their immediate goals and enable them to do their own jobs). Alternatively, a "smart" factory in which most processes are automated would likely appear system-focused to virtually anyone who actually interacts with it up close, including human workers.

People-focused. "People-focus" in smart environments fits the description from Das and Cook (2006): an environment with this focus is service-oriented, and its primary goal is to serve its human inhabitants and support the activities and goals of humans. Humans in a people-focused environment interact with technology to meet their own needs. Without regular interactions with people, the environment has little purpose.[1]

System-focused. In contrast to the user orientation of people-focused environments, system-focused environments are more concerned with their own upkeep. They "have their own objectives", so to speak, and continue to serve those objectives when not

[1]We choose the term "people-focused" here instead of "user-focused" because the latter implies that users are actively engaged with a technology and engaged by choice. In smart environments, inhabitants may be regular users but cannot be assumed to be users by default at all times simply because they are in a computation-heavy space.

occupied by people, and with or without constant input from people. Human interaction with agents in these environments will serve the goals that the environment was built to serve (including continued habitability) and will be less common than in people-focused systems. The human experience—in terms of aesthetics, comfort, and usability—will not be the main priority of the environment's existence, but still needs to be prioritized in its design.

Adaptability

Future smart homes will comprise early adopters, tinkerers, hackers, and people who resist new technologies and never want to look under the hood. In a more specialized environment like a research base or a space habitat, multiple human team members will have various non-overlapping roles, skill sets, expertise, and personal preferences, and their interactions with advanced technologies in the space will vary with these characteristics. Additionally, individual users' needs and desires for smartness in their homes will change over time. Adding more computation to living spaces requires a sensitive approach: researchers have explored how end users might "upcycle" home objects by adding lightweight modifications that do not detract or distract from the home as it stood before being smart (Williams et al., 2020), or add tags and trigger-actions to everyday objects to program them with desired "smart" behaviors (Bellucci et al., 2019). How might smart environments support the different (and evolving) characteristics of multiple users, the assimilation of new technologies, and flexible use?

Evolving. An environment that evolves can be subject to frequent additions and changes by different stakeholders, including those who create it, who oversee its operations, or who inhabit it. An example of this is a developing smart home "ecosystem": as producers release new devices and users purchase them, they meld into the existing environment and augment its capabilities. As the devices are added, the home can "become smarter" over time in the sense that the mental model of its capabilities, and perhaps its roles, broadens (e.g., it gains the ability to help a user with a chore that previously had to be done manually, stops helping a user with a chore that the user becomes better able to do without assistance from the environment, or achieves better efficiency with a task). Different inhabitants can choose the degree to which they want to interact with or even expand the autonomy (e.g., one member of the household may always use the automated door lock and add sensors or tags to her backpack and bicycle so her home lets her in when she arrives, while another prefers a physical key), and guests may never even notice it. This type of environment can also start off in a state of minimal smartness and evolve to possess a great deal of smartness through its own planned evolution and/or user-driven additions and modifications. Positioning *evolution* as an express goal can make way for an environment that does not just allow flexible use of systems, but is "designed for appropriation" (Dix, 2007).

Static. A static environment has little opportunity for frequent changes. Though upgrades that allow the environment as a whole to adapt to its users may be made, these are more likely to be provoked by someone in a supervisory role who maintains the environment (e.g., a building manager for a smart office) or require intervention from a technical expert (e.g., a programmer who can add major changes to a service) than be improvised by the environment's inhabitants. A static environment is also less able to adapt to users; instead, users need to adapt to it. This includes not only the extent to which they rely on autonomy and the interactions they do and do not have, but also the role of the environment: it cannot flexibly shift from being mentally modeled as a tool to a peer to an interactive agent. Some environments will have to be static in order to be built to last for months or years without major manual upgrades (e.g., a space habitat). As such, crucial software and hardware should be especially stable and reliable, and it should be included in environments with consideration of the fact that they may become legacy technologies.

Necessity of "intelligence"

Another critical distinction to draw is that between environments that are smart for the sake of convenience, usability, or elegance and environments that have smartness as a necessary part of their existence. How important it is for an environment to have and maintain its intelligence can affect the way people form mental models of, develop trust in, or learn about what is happening in the environment. In a smart home that can still function as a home without autonomy, learning and trust calibration can afford to be cautious and deliberate processes; that is, they can be done slowly for the sake of being done "right". Even if learning is erroneous and trust is miscalibrated, this sort of failure will not have disastrous effects. People may realize the relative triviality of developing trust and understanding in the technology; if they do, the design of the environment itself will have to somehow motivate these processes. In contrast, in an autonomous submarine, appropriate trust, fast learning, and accurate mental models are more critical and have less time to develop. In this case, people's baseline levels of motivation to engage in learning and impression formation are likely to be higher simply because of how mission-critical these processes are. As such, the environment itself may not need to be quite as attuned to actively facilitate learning.

Supplemental intelligence. Supplemental intelligence adds to the user experience for anyone interacting within the environment, but is not a non-negotiable feature of the environment. For example, a smart office building may have a connected set of sensors that monitor the inventory of the kitchen, robots that deliver food to employees, lobby information kiosks that assist visitors with navigation, and cameras that track their progress to their destination. While this might be an extremely useful and desirable set of services, the office building would still be an office building without them. Smart technologies could be removed and the environment would still be usable, and failures of intelligence present low risk.

Essential intelligence. This type of environment's survival relies on the aspects of it that give rise to its smartness. For example, in a smart space habitat that must self-sustain and support life on board, a downgrading or failure of smart aspects makes it impossible for the environment to achieve any of its goals. In a healthcare facility that is designed specifically for the easy movement and storage of assistive robots that perform critical tasks, removing the robots would undermine the facility's very existence. Failures of smart aspects of an environment of this type present high risk, and the environment would not be able to serve its purpose without its smartness. If intelligence can be conceived as a *feature* of a space when it is supplemental, then when it is essential, it is a *foundation*.

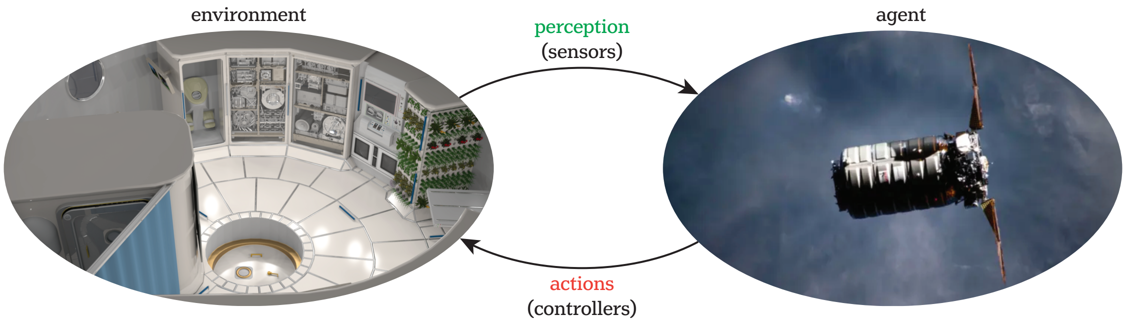

FIGURE 6.2: A smart space exploration environment as an intelligent agent. This model is modified from that proposed in Das and Cook, 2006. Left: An illustration of a smart space habitat. Right: A Cygnus cargo module, one possible design for a Lunar Gateway habitation module. Images from NASA.

Handling of attention

In some environments, inhabitants will need to direct requests to a particular device, robot, agent, or interface to achieve something. In others, sensors will be so ambient that the mental model of the environment will include the perception that anything can be done from anywhere. Smart environments can also contain some aspects that are ambient and some aspects that are directed. For example, an environment might have a temperature control system that can be accessed by voice from any number of ubiquitous microphones or by gesture through ubiquitous cameras, but it might have a lighting system that can only be controlled by interacting with a specific panel in a specific room.

Directed. Inhabitants have the perception that they must direct their attention somewhere to give commands. There may be a single focal point of interaction, such as a central voice UI that is accessed through a dedicated microphone. There could also be multiple focal points of interaction, such as different robots that need to be approached for status updates on different maintenance tasks.

Ambient. Users of the environment can direct their attention to the entire environment, or do not have to direct it at all, to achieve goals (e.g., issuing a command to turn on the lights "into the ether").

The five lenses discussed here are not the only ones through which user experience in smart environments can and should be considered. While this thesis argues that new theoretical perspectives on human-smart environment interaction—such as this one—are necessary, it also argues that it is highly unlikely that *these* five lenses are sufficient to underlie *all* of the design comparisons that could be made about different kinds of human-centered smart environments. Rather, these lenses are the basis for a new way of thinking about the smart environment as a unit of analysis for research and design of human-system interactions. They also set the state for empirical work comparing combinations of lens-related features of smart environments. The rest of this chapter describes one such study.

6.2 Study: Conversational agents in smart space habitats

In an online study, we examined questions of agent presentation in a future smart environment. Our research questions focused on mental models, task performance variables, and social perceptions of agents and environments.

- RQ1: How does agent affiliation (with users and/or domains) and the narrative perspective an agent uses in dialogue impact the way people mentally model a smart environment?
- RQ2: How do the affiliation and narrative perspective of a conversational agent embedded in a smart environment impact workload, trust, and social perceptions of the agent and the environment?

Section 6.1.1 relies heavily on the comparison of a smart home (usually "people-focused", "emergent", "evolving", and with "supplemental intelligence") to a future space habitat (likely more "system-focused", "designed", "static", and with "essential intelligence") in highlighting the ways in which smart environments can be different from each other. To better inform this discussion using the same representative examples, we chose to situate the above research questions in a simulated future space habitat. A future space habitat serves as a representative example of a task-focused, *designed*, collaborative environments with *essential* intelligence: astronauts' collaborations with intelligent systems will include interactions with individual robots (e.g., Astrobee on the ISS) as well as exchanges of information and data with disembodied AIs (e.g., when an operator and an AI system collaborate to control a remote exploration robot). In some cases, an intelligent system may include both individual robots and supervisory or portable disembodied AIs. In addition to the basic research aims of this dissertation, this aspect of our work applies directly to NASA initiatives surrounding human interaction with integrated systems. Specifically, it reveals knowledge about how to imbue such systems with interaction capabilities that support intuitive and fluent communication, task coordination, trust and comfort with the system, appropriate social interactions among team members.

The space habitat context is also well-suited to an online study with a general population sample: Because human-agent and human-robot interactions in space are so prevalent in science fiction, we anticipated that many people would be familiar with the idea of conversational AIs supporting astronauts' tasks, and therefore easily accept and understand the multi-agent, multi-task domain scenario despite its complexity. (This also means, of course, that people may have preconceived notions about the agent presentations we tested; this is discussed further in Section 6.6.) We also chose and designed the space narrative to keep the study engaging enough to prevent confounding effects of boredom and fatigue given that it entailed both repetitive tasks and multi-step logic puzzles and took place in the browser.

6.2.1 Manipulations

Motivated by the lenses to smart environment interaction and the framing of smart space habitats as agents, we manipulated two variables:

Agent affiliation: How many agents are present, and what the purpose and knowledge of each agent is.

- Singular: A single conversational agent handles all tasks and interactions for the entire habitat.
- User: Each team member has their own conversational agent that is capable of supervising/guiding/assisting with all of their tasks and activities.
- Domain: Multiple conversational agents each have one domain of expertise (e.g., experiments, crew health & nutrition, maintenance) and interact with all team members.

System Narrative Perspective: Inspired by findings from Bejarano et al., 2022, we also manipulated whether the agent uses first- or third-person language in reference to parts of the habitat and actions that will be taken. Through this manipulation, we sought to probe mental models about two aspects of the human-agent interaction in this scenario: (1) whether the agent is "ambient" (can be accessed from anywhere) or "embodied" (needs to be accessed from specific places), and (2) whether changes in conversational design affect the degree to which people anthropomorphize, trust, and project social attributes onto agents in smart environments. These topics evoke the *handling of attention* lens from Section 6.1.1, which pertains to how users need to direct their speech and attention in order to interact with the agent(s) and the social roles that agents play.

- First-person: The main agent refers to various parts of the habitat and actions using the first-person perspective.
- Third-person: The main agent refers to various parts of the habitat and actions in the third-person perspective.

Agents interacted with study participants via a text chat interface. The agent embodiment and expertise conditions determined how many agents were present in the chat, what they provided information about, and how they used language to refer to the hardware within the habitat[2]. See Figure 6.3 for a sample exchange between a participant and the chatbot(s) in the *First, User* condition vs. in the *Third, Singular* condition.

6.3 Task

6.3.1 Summary

We designed a new "Space Habitat Task" for this study, comprised of a puzzle game to be played in a browser. The game consisted of six logic puzzles (subtasks) of varying complexity, all of which were framed with narrative related to a space mission. In order to solve each puzzle, the participant had to ask the agent(s) for various numeric values and, at times, for contextual information that would help them make sense of the puzzle's instructions. Successful completion of each puzzle was required to advance to the next puzzle. The six puzzles were categorized into two phases with three puzzles each. Phase A: 3D Print a Part and Test (1–set print settings, 2–replace extruder nozzle, 3–test

[2]See Bejarano et al., 2022; Williams et al., 2021; Jackson et al., 2021 for discussions of how the linguistic cues of *self-* and *other-reference* to parts of one's own "body" or others' "bodies" can serve as "identity observables" that influence mental models of agent and robot identity.

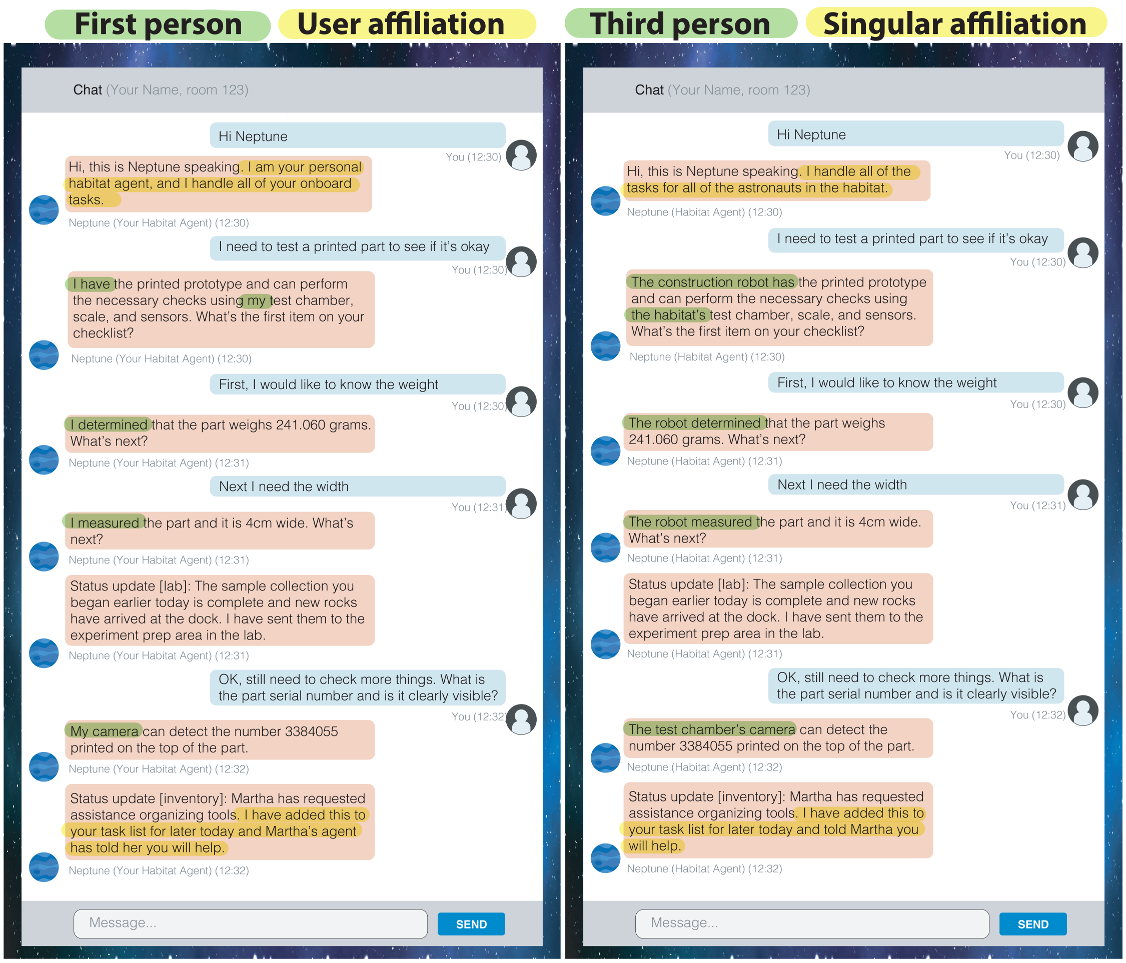

FIGURE 6.3: Two sample chat exchanges between the participant and the main agent, Neptune. In both exchanges, the participant asks Neptune for help completing a checklist involving information about a 3D-printed part (Task 1.3 in Table 6.4.2). On the left, Neptune speaks in First Person and is User-affiliated. On the right, Neptune speaks in Third Person and is Singular (not affiliated with any user or domain). Green highlights show the difference between First and Third Person dialogue; yellow highlights show the differences in the way Neptune introduces itself and talks about the participant's teammate ("Martha") in the User condition vs. in the Singular condition.

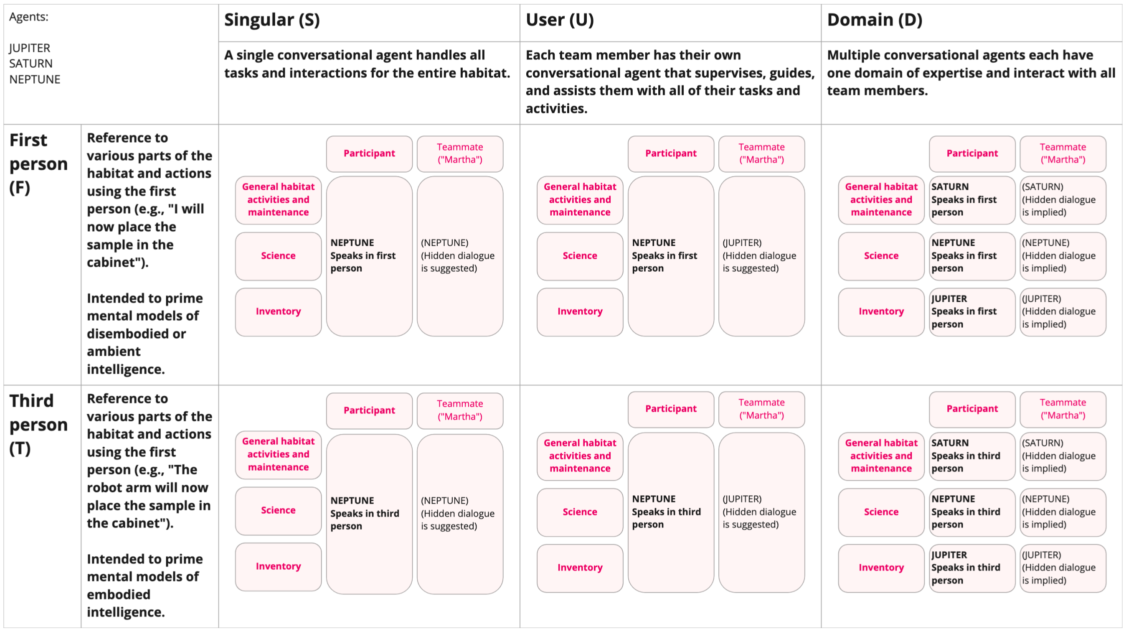

FIGURE 6.4: Condition table for the 6 agent Affiliation and System Narrative Perspective conditions.

part) and Phase B: Perform Maintenance Checks for Module Move (4–calibrate robotic arm, 5–inspect module, 6–prepare port for attaching). Table 6.4.2 shows the structure of the Space Habitat Task and communicates more details about each puzzle.

The Affiliation manipulation was signalled in two ways. First, before beginning the task, the participant clicked through several introductory screens that contained text blurbs to set up the space mission narrative, explain the puzzle structure, and introduce them to the chatbot(s). With a Singular agent, the screen about the chatbot interaction was titled "The Habitat Agent". The text explained that Neptune, the habitat's conversational agent, worked with all of the astronauts on all of their tasks. With User-affiliated agents, the screen was titled "Your Habitat Agent". The text explained that Neptune was the participant's personal habitat support agent and would help them with their tasks while their teammate, Martha, interacted with her own agent, Jupiter, to receive support for her tasks. With Domain-affiliated agents, the screen was titled "The Habitat Agents". The text explained that the participant would interact with the agent dedicated to general habitat activities and maintenance, Neptune, while two other agents, Jupiter and Saturn, provided information about science experiments and inventory, respectively.

Second, at semi-random times (on the order of every few minutes), one or more chatbots sent updates about activities that were happening in other parts of the habitat. With the single-agent conditions (Singular and User), all such updates came from the one chatbot, Neptune. With User-affiliated agents, all updates came from Neptune. With Domain-affiliated agents, updates about an ongoing *science* experiment came from a *science*-specific agent, Jupiter, while updates about *inventory* came from an *inventory*-specific agent, Saturn. To incentivize paying attention to these updates, the instructions encouraged the participant to take notes about these updates and implied that such notes would be needed later (though in reality, they were not).

Chatbots used either First Person or Third Person language to refer to robots, sensors, and other hardware within the habitat according to the System Narrative Perspective manipulation.

6.3.2 Requirements

We developed the study task to:

- Be serial and trial-based (i.e., several semi-related tasks in succession). We achieved this through an onboarding task followed by six puzzles (i.e., tasks) related to 3D printing and habitat module reconfiguration. This structure allowed participants to experience a training period for learning the interface, maintain moderate cognitive load, and base their impressions of the agent(s) based on an overall "narrative arc" of the interaction.
- Lend itself to asymmetric distribution of information across multiple channels. We achieved this by placing some information needed to complete the puzzles on the screen for participants to find and read, embedding other necessary information in the chatbot dialogue so that participants had to ask for it. We also designed any "physical" steps and the final step of each puzzle as browser interactions by the participant such that moving onto the next task always required typing a word or manipulating a control and then clicking a button. This way, participants were guided by the specifications of each task in doing computations and providing inputs while relying on the chat agent(s) for task support, as in real-world human-agent collaboration.
- Be easily learnable and understandable by the general public. We achieved this by keeping the subtasks lightweight and nontechnical such that the entire task was reminiscent of a digital escape room. This kept participants engaged and the level of challenge appropriately calibrated.

6.3.3 Narrative

The narrative was as follows:

> *You are one of two astronauts working on the Lunar Gateway habitat, a space station in lunar orbit that has numerous smart systems. The Gateway consists of multiple modules (or "rooms"—living area, laboratory, work areas, storage areas, etc.), which can be repositioned into different configurations.*
>
> *You and your crewmate, Martha, work with the habitat on your various tasks. As with all space missions, there are many tasks to be done and a tight schedule to keep. Moreover, because of mission constraints, tasks have to be performed concurrently.*
>
> *Today, you will work with the habitat on two tasks. The first task involves supervising the manufacturing and distribution of parts for a Lunar Outpost on the Moon. The habitat 3D prints the parts and performs some tests, and you track its progress, check its work, and perform other tests. Then, you work with the habitat to deploy the parts to the lunar surface.*

The second task is to perform maintenance checks of one of the habitat's modules, which will soon be detached and relocated. The purpose of these checks is to make sure that everything is ready prior to the relocation. You will perform these checks together with the habitat.

In addition to the two main tasks you are working on, you and your crewmate each have other tasks that you are working on in the background. You will receive updates from the habitat about these other tasks. For you, these other tasks involve processing samples that have been collected from the lunar surface. Meanwhile, Martha is responsible for taking inventory of equipment and supplies in another part of the Gateway. You will take notes on what you learn from these updates as you receive them.

The scenario continued with one of three descriptions (depending on the affiliation condition) of the habitat agent(s). For example, the text from the Domain condition was as follows:

To communicate with the habitat during your mission, you will interact with the habitat's conversational agents.

Note: "Conversational agents" are artificial intelligences (AI) that interact with people through language. Examples of these are Apple's Siri, Amazon's Alexa, and the Google Assistant.

For this task, the habitat conversational agents will use text chat to interact with you. The conversational agents are called Neptune, Jupiter, and Saturn. Neptune has insight into general habitat activities and maintenance, and will work with you to do your 3D printing and module maintenance tasks. Jupiter has insight into your science experiments, and will update you on the science tasks that you are working on in the background. Saturn's domain is inventory, so Saturn will update you about the activities and progress of your crew mate, Martha, as she deals with inventory.

Neptune, Jupiter, and Saturn handle tasks in their respective domains for all astronauts in the habitat—any agent can interact with any astronaut about questions and tasks in the agent's domain. Together, they can answer questions that you and Martha have while you're working, send updates, and help each of you to accomplish your tasks.

In the Singular affiliation condition, Neptune was described as having insight into and being able to work in all task domains (general habitat activities and maintenance, science, and inventory) and as being the agent for all astronauts. In the User affiliation condition, Neptune was described as being the participant's dedicated agent—able to assist *them* with tasks in all domains—and Jupiter was described as being Martha's dedicated agent, able to help *Martha* with tasks in all domains.

6.4 Study method

We recruited participants through the online research platform Prolific.co. Prolific.co is known for long-form surveys and other research tasks that ask participants to engage

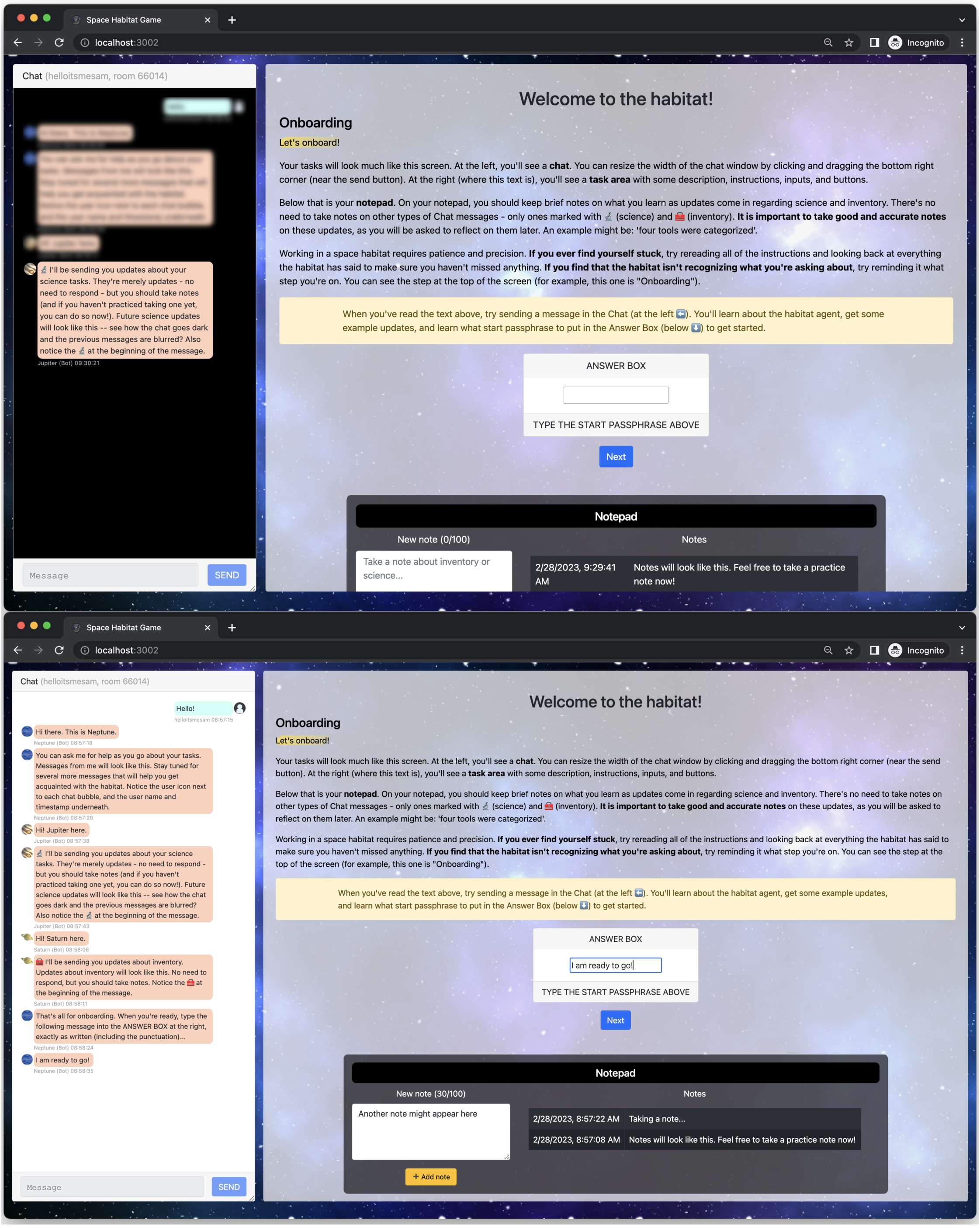

FIGURE 6.5: Two screenshots that illustrate the "onboarding" phase of the study. The chat was a resizeable panel on the left. The task area took up the remainder of the window. The introductory text provided task details and described how to use the study interface. Top: When updates about Science and Inventory came in, the chat area temporarily turned black to draw the user's attention to the update and prevent them from sending other messages for several seconds. Bottom: In order to proceed to the task, the user had to enter a passphrase given by the agent(s) at the end of the introductory sequence of messages. This screenshot shows the state of the "onboarding" phase after the agent interaction has ended, just before the user hits "Next" and proceeds to the first puzzle.

intellectually and share their opinions. It was therefore better suited to recruitment for this study than other crowdwork and online participant pools (e.g., Amazon Mechanical Turk) on which workers are more used to seeing smaller, less-involved tasks (like image labeling and A/B webpage comparisons). It has also been shown to generate higher-quality data than Mechanical Turk (Eyal et al., 2021).

6.4.1 Participants

121 people successfully completed the final version of the study. Participants were between 18 and 73 years old (mean: 36, median: 34). 45 self-identified as female, 71 as male, 3 as nonbinary, 1 as transgender, and 1 did not identify a gender. Their occupational backgrounds were varied and included accounting, freelance, graphic design, computing, IT, student, and unemployed. They used computers extremely frequently (mean: 6.9 and median: 7.0 on a 1-7 Likert-type scale), used AI assistants relatively rarely (mean: 3.0 and median: 2.0) and used robots very rarely (mean: 1.5 and median: 1.0). 22 were in Canada while participating in the study and 99 were in the U.S. 11 reported owning robots (almost all Roombas) and 90 reported owning smart assistants (mostly Apple's Siri and Amazon's Alexa). The median task duration was 50 minutes (min: 16 minutes, max: 115 minutes, mean: 54 minutes)[3] and participants were paid $10.00 each[4].

6.4.2 Study procedure

In order to be eligible for the study, Prolific users had to be 18 years old or older, be based in the United States or Canada, have a minimum approval rating of 90% for previous studies, and not have participated in any earlier versions of this study (including any of the pilots described in Section 6.4.4).[5] Eligible participants discovered the study on the list of Prolific's list of available studies. Upon electing to participate, they were directed to a Qualtrics questionnaire where they could provide informed consent.

The final page of the Qualtrics pre-questionnaire provided participants with a User ID, a Room ID, and a password to use to log in to the study website and associate their pre- and post-task questionnaire responses. It also asked them to make sure to save this information for use later in the study. Participants then clicked a hyperlink to be redirected from the pre-questionnaire to the website where the study was hosted.

[3]Duration data includes participants who gave up on the task; these participants were still paid, but their study data were excluded from analysis.

[4]Based on pilots, we expected the study to take 30-45 minutes, which would make for an hourly rate of $13.33-$20.00 USD. The range of study durations was large, which means that the hourly rate actually varied greatly across participants. The fact that we could only vaguely predict how difficult the tasks would be or how long the study would take for different people was the reason we designed a very prominent "Give Up" button into the Space Habitat Game and told participants multiple times that giving up would not affect their pay. Indeed, a number of people did give up after spending more time on the study than they wanted to, and were still paid the full amount. It is also common for Prolific participants to start a task, leave it, and return to it (within the maximum allotted time—in this case, 115 minutes), and we believe some participants did this. In retrospect, it would have been good to have a higher compensation amount; however, based on the number of free-response survey comments about how enjoyable the task was and the fact that no comments mentioned unfair pay, we do not believe participants saw the hourly rate as a problem or felt unfairly compensated.

[5]These eligibility criteria applied to all versions of the study, including the pilots described in Section 6.4.4.

Upon landing on the study page, participants logged in with the unique username and password generated for them during the pre-questionnaire, the universal password provided by the pre-questionnaire, and a username that they were asked to generate[6]. After successfully logging in, they clicked through a series of introductory screens describing the study scenario and the task (Section 6.3.3). They then completed the six puzzle-like study tasks with the help of a chatbot that was integrated into the study system. The task is described in detail in Section 6.3 and the chatbot and study system implementation are described in detail in Section 6.4.3.

When the last task had been completed, a screen explained that this portion had ended and asked participants to follow a hyperlink to a post-task questionnaire on Qualtrics. Upon finishing the post-task questionnaire, participants received a study completion code and were redirected back to Prolific, where they could enter the code to receive payment.

The study was approved by Carnegie Mellon University's Institutional Review Board.

6.4.3 Technical implementation

The study was conducted remotely online, with a browser game and text chat interface simulating interaction with one or more agents in a module of a space habitat (see Figure 6.6). The chat included both a functioning chatbot implemented using Google's Dialogflow (which handled the majority of messages sent by participants) and hard-coded, timed messages that appeared as though they were coming from chatbots, but were predetermined (which handled notifications about tasks in other domains/unrelated to the immediate task).

Chat interface and bots

We used socket.io for the chat interface. The chat element was designed to scale for multiple simultaneous multi-user chats. Though the study we report on here involved single-user interaction with an agent, we designed the system this way to preserve the possibility of conducting this study with groups (e.g., to examine User-affiliated agents with real teammates instead of the fictional "Martha").

For the Neptune agent, we developed two chatbots using Google Dialogflow. The two bots were exactly the same except that one used a First Person perspective to talk about habitat hardware while the other used Third Person perspective. Though the conversation topics were specific enough that the inputs could have been "hard-coded", we opted for a system that could leverage machine learning rather than relying solely on a rule-based chat system for a few reasons: First, the task was complicated, and we knew we would not be able to enumerate all of the possible user inputs that could progress the task. Second, it is well-known that people often try to test the limits of socially interactive systems when they first interact with them (Salvini et al., 2010; Bršćić et al., 2015; Keijsers and Bartneck, 2018; Sciuto et al., 2018). We therefore expected users to experiment with the capabilities of the chatbot, and wanted to minimize the possibility of users (intentionally or unintentionally) entering conversation loops in doing so. Third,

[6]Participants were asked not to include real names or any other potentially identifiable information in their usernames.

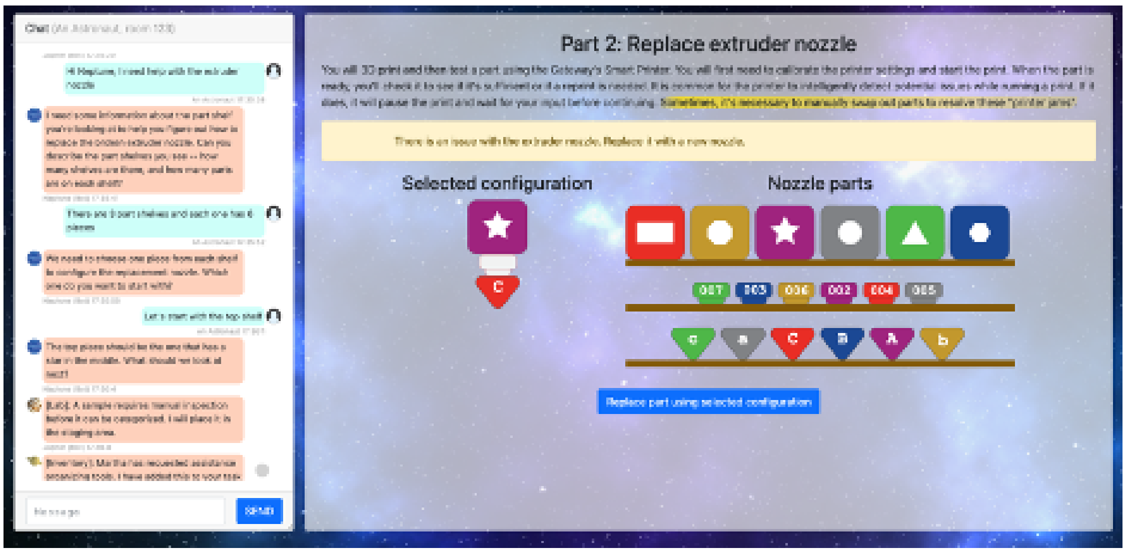

FIGURE 6.6: A screenshot simulating a participant using the browser interface for the agent presentation in smart space habitats study in the First Person (System Narrative Perspective) and Domain (Affiliation) condition. At the left, the participant requests information about the parts to be selected and the agent Neptune answers. The other two agents, Jupiter and Saturn, provide information about ongoing science experiments and inventory tasks, respectively. (Messages from Jupiter and Saturn are shown coming in in rapid succession for the purpose of illustrating the concept to test the manipulation; in the real study, there is more time between notifications.) The task shown is one of six tasks, each of which took about 1 to 5 minutes.

the learning capabilities of Dialogflow and the storage capabilities of Google Cloud allowed our agents to improve over time and were helpful for piloting, troubleshooting, gaining context about anomalous study sessions.

All of the "updates"—messages related to the secondary task domains of science and inventory—were hard-coded. This meant that in the Domain-affiliated agents condition, the "domain agents" (Jupiter and Saturn) did not actually exist outside the scope of the website code, and that in the User-affiliated and Singular conditions, a few messages from "Neptune" came from the interface rather than the actual Neptune agent. The updates appeared in the same chat window as the chat between the participant and Neptune and looked the same in visual design and conversational style as the messages from the actual chatbot.

Each of the six update messages was sent between 5 and 40 minutes after the start of the first puzzle. The ordering and exact timing of updates was random, and was determined automatically and on-line for each session after the user had clicked through the introductory screens and completed the onboarding exercise, just before they began the first puzzle.

As described in Section 6.3.1, participants were asked to take notes on the updates. Additionally, each time an update came in, the chat window turned from white to black, blurred all previous messages, and disabled (greyed out) the message box for a period of time ("reading time") corresponding to the number of words in the update $((\text{number of words} \times 300ms) + 2s + 1.5s)$[7] (see the top panel in Figure 6.5).

Task interface

We designed the task interface as a custom web app built using ReactJS, hosted with Node.js, and styled with Bootstrap. To keep the overall structure simple and ensure that participants proceeded through the instructions and tasks in the correct order, the website had only one page. Each puzzle was designed as a react component with state variables for the current guesses and a Submit button. When the guesses matched the correct answers (which were stored in a dictionary object in the overarching App component) and the Submit button was pressed, the next puzzle was rendered.

The study website was hosted on a CMU server, and was publicly discoverable as spacehabstudy.com. To proceed beyond the initial welcome screen, users had to enter a Participant ID, Room ID, and password provided by the Qualtrics pre-survey, as well as assign themselves a username for the chat. Information about the Affiliation and System Narrative Perspective conditions was encoded in the Room ID. The proper Affiliation introductory screen and agent names and icons were rendered (see Section 6.3.1) and the corresponding System Narrative Perspective chatbot API was accessed (see Section 6.4.3) accordingly. The combination of specific formatting requirements for the Participant and Room IDs and a single universal password prevented people from playing the Space Habitat Game if they somehow stumbled upon the website but were not enrolled in the study.

[7] We experimented with the reading time algorithm during pilots, eventually converging on the word count multiplier of 300ms.

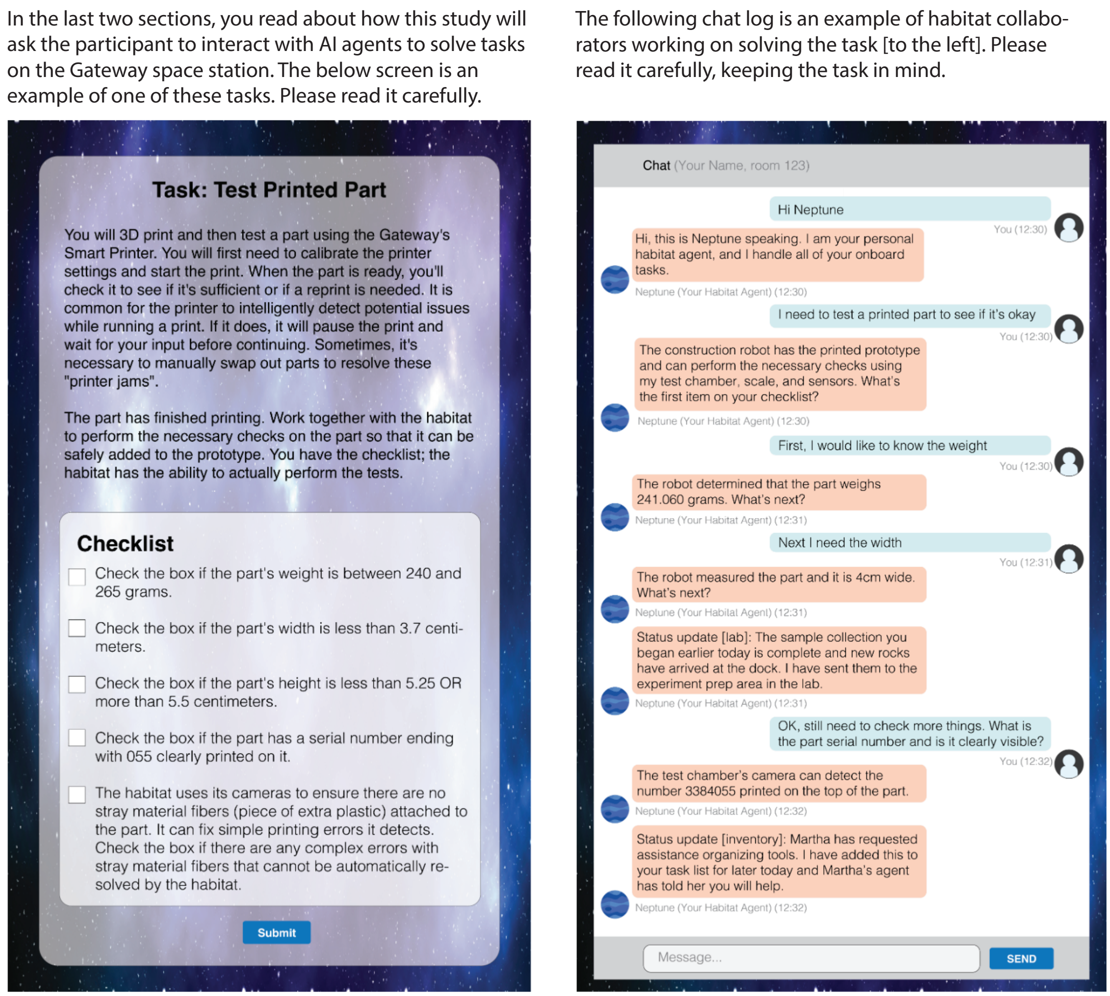

FIGURE 6.7: Screenshots from the pilot study.

6.4.4 Pilot studies

We used several rounds of piloting to refine the study scenario, calibrate the tasks to an appropriate level of difficulty, evaluate the interpretability of the chatbot's responses and develop the conversation design, and identify and troubleshoot bugs in the study interface code.

In a small initial study, we tested materials for comprehensibility with 18 Prolific participants. As with the main study, we used a factorial design for the pilot: 6 participants saw screenshots from each of the the Domain, User, and Singular agent conditions; 9 participants saw screenshots from each of the First- and Third-person conditions. The study first asked participants to closely examine the two initial text block screens, which introduced the study scenario and began the narrative. Then, it displayed a screenshot of one of the six tasks and a mock conversation between a user and the chatbot (whose mock responses were written according to one of the six conditions) about the task (Figure 6.7). To check the agent affiliation manipulation, we asked multiple-choice comprehension questions about the roles of the agents (e.g., "If you were the human in the chat

("You") and you wanted to ask the habitat to check something in the inventory, which agent would you ask?"). We also asked open-ended questions about what people found confusing about the task, what they liked about the task, and their general reflections on the task. This pilot study took about 10 minutes to complete, and participants were paid USD $5.00.

Then, we tested the study interface (including the chatbot interactions) for usability with a handful of colleagues. Once we had resolved the usability issues we could identify on our own, we ran several small-scale pilot studies with the actual system and the full post-task questionnaire. After each study, we reviewed user interaction data and questionnaire responses: we looked for indications of task components that were too hard (e.g., subtasks on which users spent a disproportionate amount of time or that caused them to give up, common wrong answers to puzzles), common triggers for Dialogflow fallback intents (i.e., chatbot inputs that were not recognized) and improperly recognized intents, and questionnaire feedback that highlighted bugs and usability issues. Through these pilots, we also identified opportunities to ask about perceptions of aspects of the study that we did not initially consider; we expanded the post-task questionnaire accordingly.

All pilot studies on Prolific had the same eligibility criteria as the main study, and all participants gave informed consent.

6.4.5 Measures

We measured performance in the form of task time and accuracy. We used the NASA-TLX (Hart and Staveland, 1988) to measure workload[8].

We had several measures for different kinds of social perceptions: we assessed perceptions of the anthropomorphism (of the agent Neptune and of the habitat) using analogical statements from Ezer's anthropomorphism instrument (Ezer, 2008); the social constructs of perceived warmth, competence, and discomfort (for both the agnet Neptune and the habitat) using the Robotic Social Attributes (RoSAS) scale (Carpinella et al., 2017); and trust (in Neptune and in the habitat) using version 1 of the Multidimensional Measure of Trust for HRI (Ullman and Malle, 2019). We included questions specific to this study about task difficulty (of each puzzle and of the task in general) and about mental models of the agent(s). We also asked open-ended questions about impressions of the task and the agent(s). Finally, we asked several demographic questions, including questions about people's familiarity with agents, robots, and puzzle games and chat interfaces similar to our study.

6.5 Results

We performed Bayesian ANOVAs using JASP (Bergh et al., 2020) to compare responses to the questionnaires by Affiliation (Singular, User, Domain) and System Narrative Perspective (First Person, Third Person) conditions. We opted for Bayesian rather than frequentist methods because of the open-ended nature of our exploration. In this research, one of our goals is to start to determine what kinds of flexibility designers have

[8]The NASA-TLX is typically administered using paper and pencil or custom-built slider UIs if done digitally. As part of this project, we created documentation for how to administer the NASA-TLX using Qualtrics: https://github.com/CMU-TBD/qualtrics-tlx

when creating conversational AI agents for complex environments vs. what principles they should stick to across contexts. In order to conclude that designers can have the flexibility to choose different system intelligence design metaphors; vary agent interaction capabilities and styles; and affiliate agents with users, domains, or neither; such choices must *not* have large effects on critical HAI outcomes. In contrast, identifying principles that *should* be followed across contexts (or in certain types of environments) requires determining what variables *do* affect those outcomes. Bayesian methods have the benefits of being able to quantify evidence for or against the existence of an effect of a model variable and to use prior distributions from existing work[9] (see Keysers, Gazzola, and Wagenmakers, 2020). In recent years, the field of HRI has seen an increase in the number of papers that use Bayesian analysis (e.g., Banisetty and Williams, 2021; McCaffrey et al., 2021; Winkle et al., 2022; Bodala et al., 2020; Bejarano et al., 2022; Wen et al., 2021) because of these advantages, because it permits the continual gathering of data until some evidence for an effect or the absence of an effect is found (Winkle et al., 2022), and because of a growing set of concerns about the limitations of heavy reliance on p-values produced by Null Hypothesis Significance Testing (NHST) in the HRI (Hoffman and Zhao, 2020; Winkle et al., 2022) and broader scientific (Kruschke, 2010; Dienes and Mclatchie, 2018; Krueger and Heck, 2017) communities.

The post-study questionnaire included questions about "*the agent*" (A) and questions about "*the habitat*" (H). "The agent" refers to the main agent, "Neptune", which was also the sole agent in the Singular condition. In the Single and User condition, Neptune handled all of the user interaction and provided all of the updates, so it was the only agent that participants saw. In the Domain condition, participants interacted with Neptune, but also saw updates from Jupiter and Saturn; because Jupiter and Saturn were peripheral agents (there was minimal exposure to them and they were not interactive), we did not ask questions about them in our questionnaire. As explained in Section 6.4.3, in all conditions, Neptune was the only actual chatbot; updates from other agents ("Jupiter" and "Saturn") in the Domain condition were hard-coded, but appeared as though they were coming from live chatbots.

In cases where (1) the distributions of perceptions about the agent and perceptions about the habitat for a specific variable are extremely similar, we report only on perceptions about the agent. Where there is a non-trivial difference in the scores, or where differences in attributions to "the agent" vs. to "the habitat" are semantically meaningful for our research questions, we report on such differences.

In Bayesian ANOVAs, evidence in favor of the absence of any main effect is also evidence against any interaction effects. We do not report on models that violate the principle of marginality—that is, models which feature interaction effects without their constituent main effects (see Nelder, 1977). Table 6.2 shows the commonly used guidelines for interpreting the strength of evidence in favor of null and alternative hypotheses using Bayesian analysis (see Lee and Wagenmakers, 2014).

[9] Our study did not use an informative prior because it was almost entirely an initial exploration; that is, there is no existing quantitative evidence from previous studies on which to base a prior. As the research topic of agent identity affiliation evolves, future analyses will be able to determine informative priors.

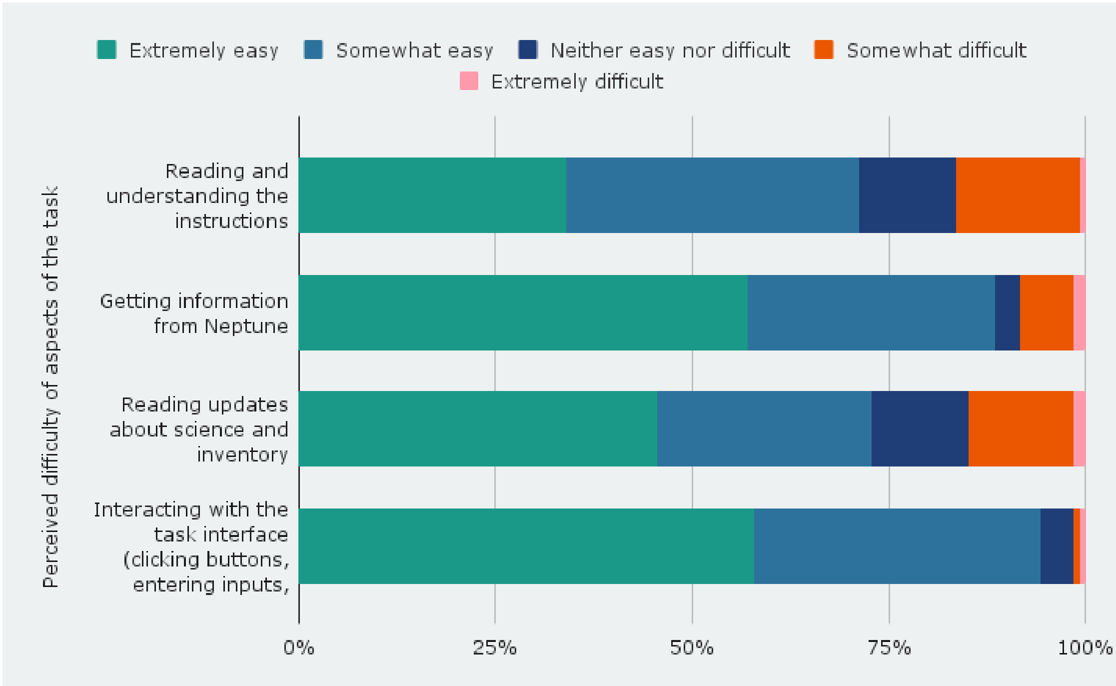

FIGURE 6.8: Chart showing how easy or difficult participants perceived various aspects of the study to be.

6.5.1 Task experience

Nearly all participants found interacting with the task interface to be "extremely easy" or "somewhat easy". Most also reported finding it easy to get information from Neptune. Participants on average found it was slightly more difficult to read and understand the instruction and to read updates about science and inventory (see Figure 6.8).

6.5.2 Manipulation check

While the pilot studies gave us confidence that participants were making sense of the task and manipulations, we wanted to confirm that participants accurately understood the affiliation conditions in the full study. We asked questions about whether Neptune was their personal interface to the habitat and whether it was Martha's interface to the habitat. The most critical manipulation check was whether or not participants understood the User affiliation. This is because in the User condition, the existence of a second agent (Jupiter) was suggested, but the agent was not actually present. Signals that participants were interacting with user-affiliated agents came only in the form of study framing (i.e., in the introductory text) and through Neptune's references to Jupiter during the dialog; therefore, there was some risk that user-affiliated agents would be indistinguishable from a singular agent. The Domain affiliation was much more obviously different from the other two affiliations because the user actually interacted with multiple agents.

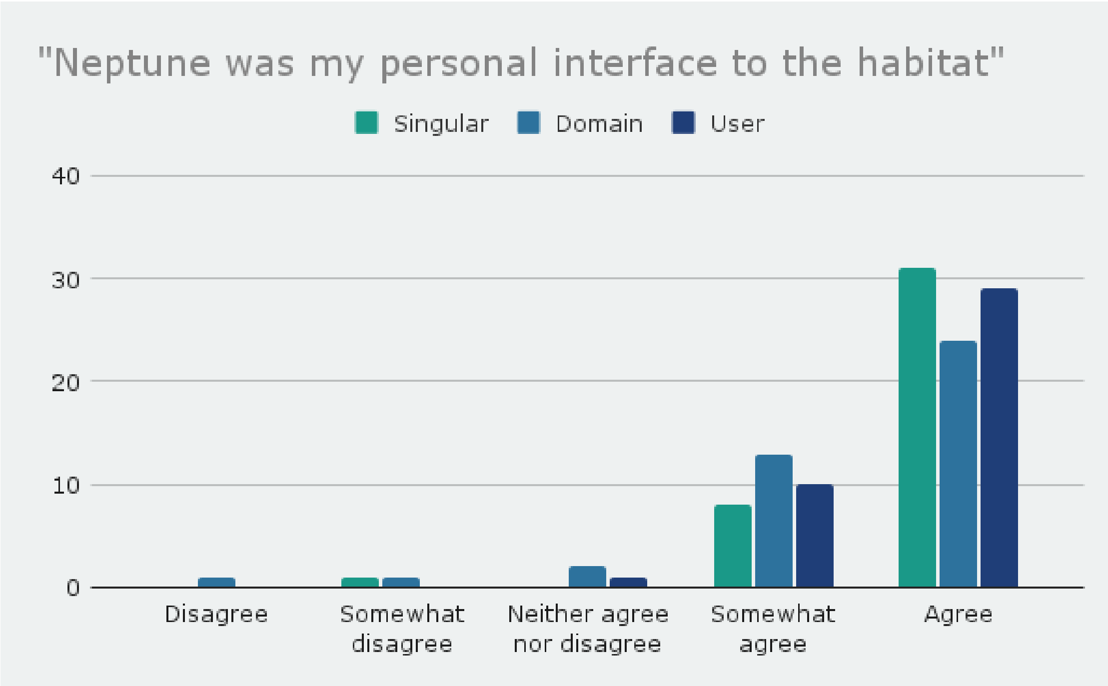

FIGURE 6.9: Perceptions of the agent as "my personal interface to the habitat" by Affiliation condition.

In all conditions, most participants strongly agreed that Neptune was their personal interface. This suggests an accurate understanding of the User conditions[10]. In the User condition, most participants strongly disagreed that Neptune was Martha's interface. In the Singular condition, most participants strongly agreed that Neptune was Martha's interface. This suggests that they understood the difference. Figures 6.9 and 6.10 summarize participants' responses to these questions.

We did not check participants' understanding of the System Narrative Perspective manipulation because first- vs. third-person references can implicitly influence perceptions without being explicitly noticeable.

6.5.3 Item reliability

We calculated Cronbach's α for the items from the RoSAS (Carpinella et al., 2017) measuring perceived competence, warmth, and discomfort for both the agent and the habitat. All scores fell between 0.80 and 0.89, which is considered good reliability. We calculated Cronbach's α for the four subscales of trust from the MDMT (Ullman and Malle, 2019): capable, reliable, ethical, and sincere. For perceptions that the agent was reliable, $\alpha = 0.72$, which is considered fair. All other factors had good reliability, with scores

[10] Arguably, this also suggests an accurate understanding of the Single and Domain conditions, where Neptune was the user's primary interface, but not necessarily their "personal" interface. Since we did not ask the same question without the word "personal", we cannot determine whether participants actually thought that Neptune was user-affiliated even in the Singular and Domain conditions, or if they understood that Neptune was not affiliated with anyone in particular but interpreted the question differently from how we intended it.

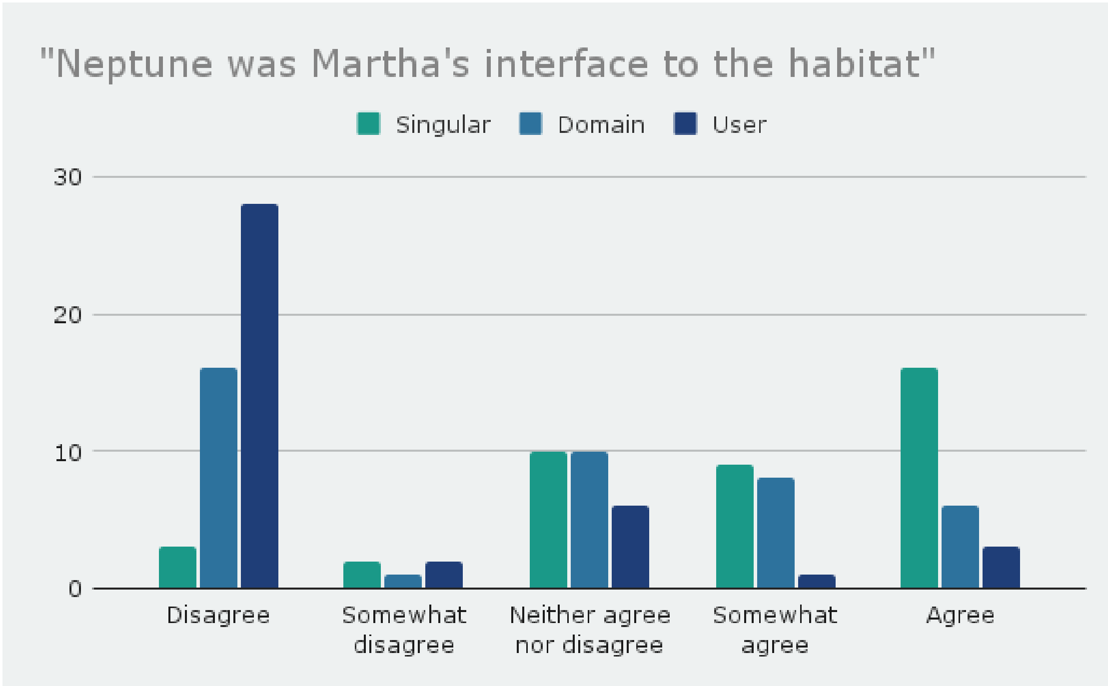

FIGURE 6.10: Perceptions of the agent as "Martha's interface to the habitat" by Affiliation condition.

ranging from 0.8 to 0.94. For agency (Kozak, Marsh, and Wegner, 2006), α scores were 0.87 and 0.89 (good) for the agent and the habitat, respectively.

6.5.4 Trust

Because of an error in our questionnaire, the trust questions were optional, and not all participants answered every question. Therefore, data from between 1 and 25 participants are missing from each of the following calculations.

The agent and the habitat were seen as generally reliable ($M_A = 5.83, SD_A = 1.17$; $M_H = 5.18, SD_H = 1.46$) and capable ($M_A = 6.01, SD_A = 1.35; M_H = 5.63, SD_H = 1.50$); distributions were left-skewed. Participants' perceptions of the agent and the habitat as ethical ($M_A = 4.80, SD_A = 1.84; M_H = 4.62, SD_H = 1.90$) and sincere ($M_A = 4.68, SD_A = 1.85; M_H = 4.43, SD_H = 1.90$) were more varied with distributions closer to normal (Figure 6.11).

Perceptions of a system as *reliable* and *capable* can be combined into a single measure of *capacity trust*. There was moderate evidence of an effect of System Narrative Perspective on capacity trust in the agent (BF = 5.02) and weak evidence of an effect of System Narrative Perspective on capacity trust in the habitat (BF = 1.99). Capacity trust was highest in the Third Person condition (on a scale of 0 to 7, $M_A = 6.19, SD_A = 0.77; M_H = 5.65, SD_H = 1.31$) than in the First Person condition ($M_A = 5.61, SD_A = 1.48; M_H = 5.03, SD = 1.59$).

Perceptions of a system as *ethical* and *sincere* can be combined into a single measure of *moral trust*. We found moderate evidence in favor of the absence of an effect of Affiliation (BF $= 0.21$) and weak evidence for the absence of an effect of System Narrative

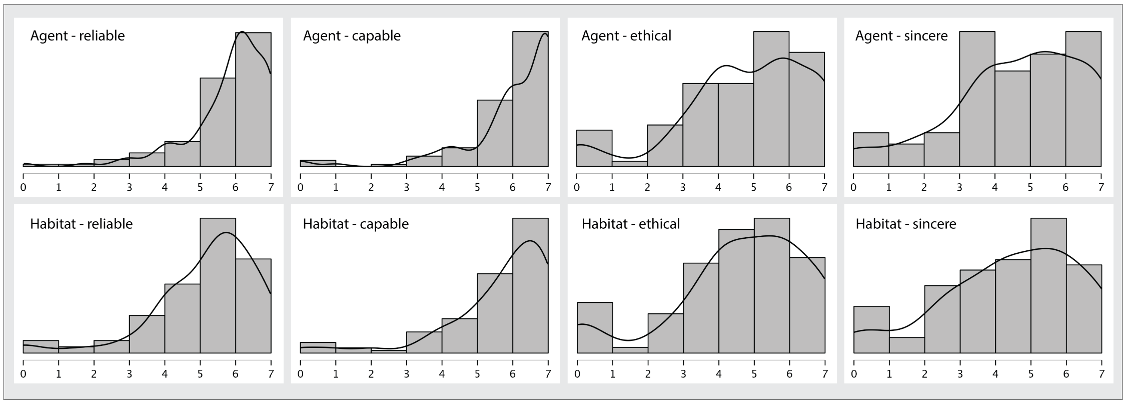

FIGURE 6.11: Distribution and density plots for the four components of trust for the agent and the habitat. Responses of "does not fit" are excluded from the figure.

Perspective (BF = 0.48) on moral trust in the agent.

6.5.5 Social attributes

Across all conditions, participants perceived both the agent and the habitat to be generally competent ($M_A = 7.71, SD_A = 1.30; M_H = 6.66, SD_H = 1.85$), not particularly warm ($M_A = 2.69, SD_A = 1.60; M_H = 2.65, SD_H = 1.59$), and not discomforting ($M_A = 1.65, SD_A = 1.0; M_H = 1.82, SD_H = 1.06$).

We found weak evidence for the absence of an effect of Affiliation and System Narrative Perspective on warmth and discomfort for both the agent and the habitat (all BFs between 0.10 and 0.49). We found strong evidence that System Narrative Perspective impacts attributions of competence to the habitat (BF = 23.57). Participants perceived the habitat to be more competent in the Third Person condition ($M = 7.19, SD = 1.51$) than in the First Person condition ($M = 6.12, SD = 2.02$). Interestingly, this was not mirrored in perceptions of the agent's competence (BF = 0.47).

6.5.6 Agency

We found strong evidence for the absence of an effect of Affiliation on mind attribution (Kozak, Marsh, and Wegner, 2006) to the agent (BF = 0.08) or to the habitat (BF = 0.09) and moderate evidence for the absence of an effect of System Narrative Perspective (agent: BF = 0.19, habitat: BF = 0.21).

6.5.7 Mental models of the agent and the habitat

Anthropomorphism instrument analogical statements

Participants perceived the agent and the habitat to be much more "like an assistant" and "like a teammate" than "like a pet" (see Figures 6.12 and 6.13). There was weak-to-strong evidence that neither manipulation affected these perceptions (all BFs between 0.09 and 0.44).

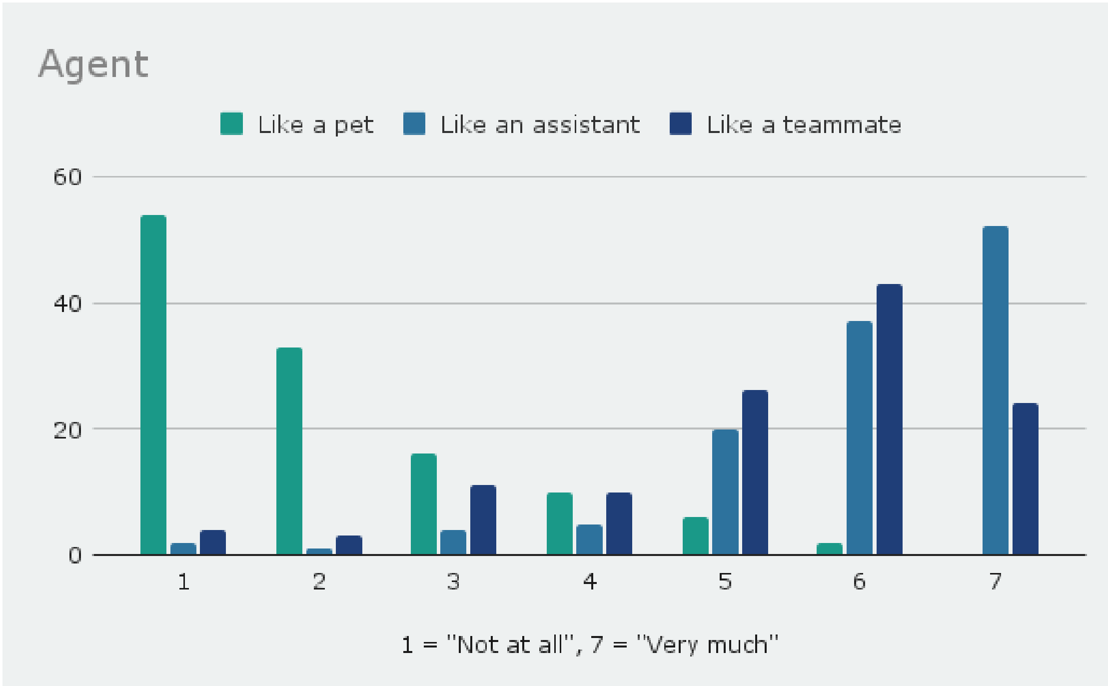

FIGURE 6.12: Perceptions of the agent as “like a pet”, “like an assistant”, and “like a teammate” (Ezer, 2008) on a scale of 1 to 7. “Like an assistant” was the dominant analogy.

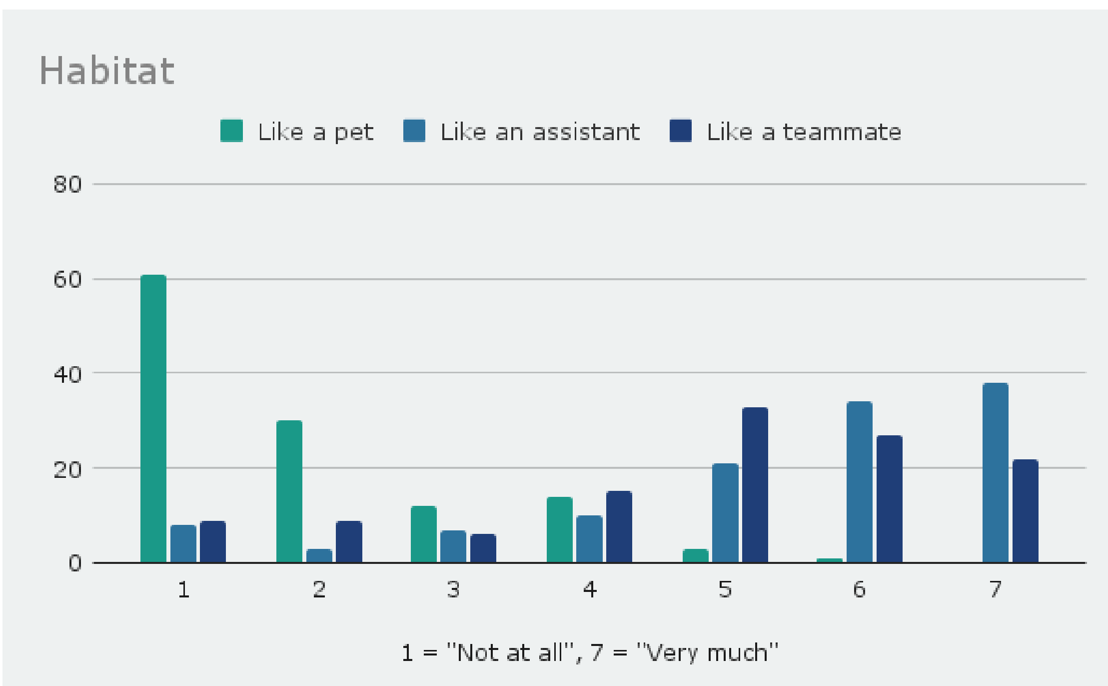

FIGURE 6.13: Perceptions of the habitat as “like a pet”, “like an assistant”, and “like a teammate” (Ezer, 2008) on a scale of 1 to 7. “Like an assistant” was the dominant analogy, but the association was not as strong for “the habitat” as it was for “the agent”.

Relationship between the agent and the habitat

We found weak evidence for an impact of affiliation on perceptions that the agent "*is* the habitat" (BF $= 1.87$). Participants agreed with this statement the most in the Singular condition (on a scale of 1 to 5, $M = 3.28, SD = 1.28$), followed by the User condition ($M = 2.78, SD = 1.48$), and then the Domain condition ($M = 2.46, SD = 1.21$). Post-hoc tests revealed moderate evidence for the difference between the Domain and Singular conditions (BF $= 8.82$) and weak evidence for the difference between the Single and User conditions (BF > 0). There was weak evidence in favor of the absence of an effect of System Narrative Perspective on this perception (BF $= 0.44$).

For perceptions that the agent "*represents* the habitat" we found weak evidence for an effect of System Narrative Perspective (BF $= 2.71$). Participants agreed with this statement more in the Third Person condition (on a scale of 1 to 5, $M = 4.13, SD = 0.76$) than in the First Person condition ($M = 3.68, SD = 1.21$). There was moderate evidence in favor of the absence of an effect of Affiliation on this perception (BF $= 0.14$).

6.5.8 Responses to updates

Overall, participants found the updates somewhat annoying ($M = 3.41, SD = 1.42$) and distracting ($M = 3.65, SD = 1.51$), which was our expectation. We found moderate evidence that neither manipulation impacted how annoying or distracting participants perceived the updates to be (all BFs between .10 and .22).

6.5.9 Workload

There was moderate evidence for an effect of Affiliation condition on the mental demand component of the TLX (BF $= 6.96$). Mental demand was highest in the Singular condition (on a scale of 0 to 100, $M = 59.15, SD = 23.07$) followed by the User condition ($M = 59.15, SD = 25.83$) and then the Domain condition ($M = 45.15, SD = 25.00$). Post hoc tests revealed strong evidence for the difference between the Domain and Singular conditions (BF $= 15.73$) and moderate evidence for the difference between the Domain and User conditions (BF $= 3.18$). There was moderate evidence for the absence of an effect of System Narrative Perspective on mental demand (BF $= 0.23$). There was strong evidence for the absence of an effect of Affiliation on the performance component of the TLX (BF $= 0.08$), and weak evidence for the absence of an effect of System Narrative Perspective (BF $= 0.46$). There was weak-to-moderate evidence for the absence of an effect of either manipulation on all other components (physical demand, temporal demand, effort, and frustration) (all BFs between 0.19 and 0.89).

6.5.10 Correlations among trust, social perceptions, and mind attribution

We used Bayesian Pearson correlations to examine relationships among different kinds of attributions (i.e., of trust, of social perceptions, and of mind attribution) to the agent and the habitat. There was moderate to strong evidence for correlations among several of these variables. These are listed in Table 6.3.

6.6 Discussion

We structure our discussion around our research questions. We also highlight limitations of this work. Throughout the section, we suggest opportunities for future research.

RQ1*: How does agent affiliation (with users and/or domains) and the narrative perspective an agent uses in dialogue impact the way people mentally model a smart environment?* In the first part of this chapter (Section 6.1.1), we discussed how smart environments might be designed differently in terms of *handling of attention*. Designers may want to create a user experience that suggests that users need to direct their attention to a single interface, to different interfaces for different tasks, or to nothing in particular. This concept is strongly intertwined with the concept of whether end-users perceive multiple intelligent systems to be "one" vs. "multiple", which is a central theme throughout this thesis. In this study, we examined the concept of "one vs. multiple" by probing people's mental models (broadly defined) of "the agent" vs. "the habitat" and examining whether they differed based on agent affiliation and narrative perspective. We found that a Singular agent—which might be associated with everything and everyone, or nothing and no one in particular—was perceived as *being* the habitat more than were Domain-affiliated agents or User-affiliated agents. A potential implication of this is that if a specific smart environment is intended to be seen as a unified entity, then user experience within it should involve a single interactive agent or intelligence. This is one way of implementing the smart environment as intelligent agent model in Figure 6.2.

We also found weak evidence that people see an agent as *representing* the environment in which it is embedded more if it uses third person pronouns to refer to hardware within that environment. The notion that one entity "represents" another suggests a distinction between the two entities (where as "is" suggests sameness). Intentionally distinguishing between a software identity and robot hardware as part of robot user experience design might be a beneficial interaction strategy in some scenarios: Williams et al. (2021) showed how people can attribute trust to a physical robot embodiment differently from how they attribute trust to an intelligence that resides in that embodiment, and Reig et al. (2021a) (Chapter 5) showed how disentangling robot identities from robot bodies might help with preservation of trust and positive perceptions of a robot system after it fails. Our work suggests that third person reference cues is one way of making such a distinction.

RQ2*: How do the affiliation and narrative perspective of a conversational agent embedded in a smart environment impact workload, trust, and social perceptions of the agent and the environment?* Our prior work (in particular the studies described in Chapters 3 4) surfaced potential concerns with designing one agent that appears to have expertise across multiple domains. Previous work found that having multiple agents discuss different tasks in the same service domain brought about confusion and little benefit (Chaves and Gerosa, 2018). In this study, our goals included exploring the relationship between those insights and better understanding how they might apply to human-agent collaboration settings. We found moderate evidence that interacting with a singular agent leads to higher mental demand than interacting with multiple agents. A possible interpretation of the combination of the insights from this study, Chaves and Gerosa (2018), Reig et al. (2020) (Chapter 3 of this thesis) and Reig et al. (2021b) (Chapter 4 of this thesis) is as follows: Having multiple agents is a beneficial design when the tasks or "jobs" are vastly different; for example, scheduling an appointment and also supporting a doctor during a physical exam (a service example from Chapter 3) or categorizing geological samples and also taking inventory of office supplies (a collaboration example from this chapter). However, when tasks or "jobs" cluster together—for example, checking a traveler in for a flight and also helping them board the plane (a service example from Chapter 4) or scheduling meals and also planning activities (a collaborative planning example from Chaves and Gerosa, 2018)—having multiple agents adds confusion and load without any benefit to trust and social perceptions, and thus only serves to complicate the interaction.

We also saw some evidence that agent affiliation with users and/or domains does not impact capacity trust (i.e., trust that an system is reliable, capable, and performs its functions as expected) or moral trust (analogous to human relational trust) in the agent. This suggests that designers can have some flexibility in designing an agent's performance of affiliation with topics and with individual people without worry that making the "wrong" choice about such performance could risk substantially negatively impacting trust. This is in line with a finding and design implication from Chapter 5, in which we found that people had diverse preferences for multi-robot system failure recovery strategies, and suggested that designers choose the strategy that was best suited to situational and/or system constraints.

It also seems that third person language cues gave people more confidence in an agent: they perceived the third person System Narrative Perspective to be more trustworthy (in terms of capacity trust) and more competent. Therefore, designers of intelligent systems that use natural language to interact with people may want to consider using a third person perspective to increase trust and perceived competence of the system. Inverting this point also reveals a design implication: Prior work has found potential risks of anthropomorphic robot design for occupational safety (Onnasch and Hildebrandt, 2021; Lee and See, 2004) and detailed the ethical dangers of people placing more trust in robots than they "should" (Coeckelbergh, 2012; Scheutz and Malle, 2018; Kok and Soh, 2020 and numerous others). Research has also shown that robots that are seen as having more animacy are less likely to be perceived as intelligent (Bartneck et al., 2009a). In this vein, it may be beneficial for natural language systems to be designed to use a *first* person perspective to *reduce* people's perceptions of trustworthiness and competence. Given recent advances in research into how robots can both assess and express their own incapability to human teammates (Kwon, Huang, and Dragan, 2018b), the same natural language system might even alternate between third person and first person perspectives depending on its confidence with a particular task or at a particular time. Conversational agents are becoming more prevalent, more capable, and more accessible, and are being used to support an increasing number of people with an increasing number and variety of tasks. As this trend continues, future research should investigate the specific effects of pronoun usage and narrative perspective on interaction outcomes.

Limitations and future work

Some limitations of this work arise from the fact that participants in this study encountered a flattened version of human-agent collaboration. Participants played the Space Habitat Game in a single browser window and with chatbots that interacted only through written text. The lack of any embodied interaction detracts somewhat from our goal of studying HAI with multi-embodiment systems.

The study narrative also brings about some limitations. First, many people may have preexisting associations with human-agent interaction in space. It is likely that the combination of the space context and our explicit focus on "one" vs. "multiple" entities made certain science fiction and media associations even more prominent in participants' minds. People may have brought such associations to the study, and it is possible they influenced their ratings about the agents and the habitat and therefore our results. Second, the tasks that participants were asked to complete in this study were inspired by real tasks that astronauts might encounter in space habitats, but were extremely oversimplified in their requirements and explanations, and did not mimic real astronauts' tasks with any temporal realism (for instance, in the study, participants were told that the 3D printed part was finished printing mere seconds after it started). In order to be meaningful for real human-system teaming in space, our findings would need to be validated with populations of experts, and through in-person studies employing more realistic scenarios and more diverse interfaces. This study also examined perceptions of user-affiliated agents with only one actual user. Future work should explore this concept with actual, rather than fictional, teammates.

Finally, there are several limitations of our measures. We probed mental models through analogical statements and free response questions. We did not have a way of validating the statements within our study, so we do not know if people interpreted the terms "is the habitat", "represents the habitat", "assistant", "pet", and "teammate" consistently, or the same way as we did. This perspective on mental models only provides a very small window into the way people perceived the relationship between the agent and the environment. Future work should conduct a deeper exploration of mental models, i.e., through interviews, think-aloud protocols with the Space Habitat Task or similar tasks, or by asking people to draw the agent and the habitat. We also relied entirely on self-report measures, which are very limited in their ability to reflect and anticipate how people truly behave and respond during interactions with autonomous systems. Future research on this topic should consider using behavioral measures of trust (e.g., how much a user intervenes with a system's autonomous behavior when intervention is not needed, a user's willingness to use a system again).

6.7 Summary and contributions

This chapter articulated a need to better understand how to design user experience in smart environments from a bird's-eye, whole-system perspective. It described five *lenses* that researchers and designers might employ in outlining requirements for smart environments, creating human-system interactions in smart environments, and studying the behavior of people and autonomous in the context of smart environments environments. Then, it reported on the results of a study that used a puzzle task developed as an online space mission game in combination with real and simulated chatbots to explore impacts of agent affiliation (with particular people or particular domains) and system narrative perspective (first vs. third person) on perceptions of an agent, the task, and the environment. Our findings reveal design implications for human-agent interaction design in smart environments and highlight areas for future research into this topic.

The work described in this chapter makes the following contributions:

- We found that a singular agent was more likely to be perceived as "being" the environment than were domain- and user-affiliated agents.
- We found that an agent that spoke in third person was more likely to be perceived as "representing" the environment.
- We found evidence that a third person language perspective can lead to increased trust and perceptions of confidence over a first person perspective.
- Our findings suggest that in this kind of setting, agent affiliation does not impact trust or social perceptions of the agent or the environment.
- We designed the "Space Habitat Task", which can be used in future studies to simulate collaboration among humans and agents in smart environments.

ID	Subtask name	Instructions	Summary of puzzle	Solution description	Participant's contribution	Agent's contribution
PHASE A: 3D Print a Part and Test						
Narrative:						
It's time to start your first task for the day. You will 3D print and then test a part using the habitat's Smart Printer. You will first need to calibrate the printer settings and start the print. When the part is ready, you'll check it to see if it's okay or if it needs to be reprinted. It is common for the habitat printer to detect problems while running a print. If there is a problem, it will pause and wait for your input before continuing. Sometimes, it's necessary to manually swap out parts to resolve these "printer jams".						
1	Set print settings	*Calibrate the printer settings and print a prototype of the part for the Lunar Habitat.*	Look at the current printer settings and adjust them so they match the required settings	Four values for speed, temperature, infill, supports input via text boxes (integer array)	Asks for and inputs the values	Provides the values
2	Replace extruder nozzle	*There is an issue with the extruder nozzle. Replace it with a new nozzle.*	When a printing issue occurs, identify and select three nozzle parts needed to create a new extruder nozzle	Three shapes ("parts"), selected by the participant from a grid of eighteen shapes (object array)	Asks the agent for information; evaluates logical statements and identifies and selects the correct parts accordingly	Describes the correct parts
3	Test part	*Complete the tests on your list. When you've performed all the checks, put the part in storage until it is ready to be attached.*	Fill out a checklist about the printed part (including information about measurements, printing errors, etc.) and enter an ID for a cabinet in which to store the part	Correct set of true/false values for six checklist items (boolean array) + a cabinet ID code (string)	Asks the agent for information; evaluates logical statements and fills in checklist accordingly; asks for and inputs cabinet ID code	Provides information about the part's measurements and print specs; provides the cabinet ID code
PHASE B: Perform Maintenance Checks for Module Move						
Narrative:						
You're done with Part 1: 3D printing and testing a part. The notes you've taken have been stored, and your notepad will refresh for Part 2. In Part 2, you will perform maintenance checks to prepare one of the habitat's modules for repositioning tomorrow. First, you will inspect and calibrate the Smart Robot Arm that you will be using tomorrow to detach and reattach the module. Then, you will inspect the module. Then, you will inspect and calibrate the new location for the module.						
ID	Subtask name	Instructions	Summary of puzzle	Solution description	Participant's contribution	Agent's contribution
4	Calibrate robotic arm	*Calibrate the Smart Robot Arm. The habitat can guide you through the calibration process.*	Select the correct values for the sensors and actuators (grip force, camera aperture, and angle of rotation of four joints) of the robot arm that will detach and reattach the module	Six numeric values provided via four knobs, one horizontal slider, and one dropdown menu (integer/string array)	Asks the agent for information; gives information about the changing position of the end effector back to the agent; rotates knobs, positions slider, and selects dropdown option	Provides values for the joint rotation, grip force, and aperture; asks for x, y, and z values of the end effector
5	Inspect module	*Perform the checks and complete the list accordingly.*	Look at a diagram of the module, fill out a checklist about the module (including information about internal pressure, usage of stowage compartments), then apply an approval code to the checklist	Diagram ID (string) + correct set of true/false values for four checklist items (boolean array) + approval code (string)	Reviews and analyzes the diagram; combines information with information requested from the agent to fill out checklist	Provides information about the pressure inside the module; provides the approval code
6	Prepare port for attaching	*Calibrate the port. Once the levers and codes are set, sign off on the settings.*	Prepare the port that the module will attach to by toggling levers and numeric codes, then sign off on the module move with an assurance signature	Correct left/right values for six levers (boolean array) + correct on/off values for five numeric codes (boolean array) + assurance signature (string)	Asks the agent for the lever positions; gives information about the resulting values back to the agent; chooses which codes to apply based on what the agent says about them; types and submits the assurance signature	Describes which levers to set to which positions; explains how the codes should be selected or left unselected; gives format for the assurance signature

TABLE 6.1: Details including the instructions, summary, solution, participant's contribution, and agent's contribution for each of the the six puzzles in the Space Habitat game.

Bayes Factor BF_{10}	Label
< 0.01	Extreme evidence for H_0
$0.01 - 0.0\overline{33}$	Very strong evidence for H_0
$0.0\overline{33} - 0.1$	Strong evidence for H_0
$0.1 - 0.\overline{33}$	Moderate evidence for H_0
$0.\overline{33} - 1$	Weak/anecdotal evidence for H_0
1	No evidence for H_1 or H_0
$1 - 3$	Weak/anecdotal evidence for H_1
$3 - 10$	Moderate evidence for H_1
$10 - 30$	Strong evidence for H_1
$30 - 100$	Very strong evidence for H_1
> 100	Extreme evidence for H_1

TABLE 6.2: Classification of Bayes factors (Lee and Wagenmakers, 2014)

			CT		**MT**		**C**		**W**		**MA**	
			A	H	A	H	A	H	A	H	A	H
CT	CA	*r*	-									
		BF	-									
	H	*r*	.61	-								
		BF	1.85e10	-								
MT	A	*r*	.64	.58	-							
		BF	9.28e10	3.10e8	-							
	H	*r*	.48	.74	.79	-						
		BF	9.48e4	3.96e16	8.30e19	-						
C	A	*r*	.31	.60	.40	.44	-					
		BF	35.78	8.21e9	1.09e3	9.10e3	-					
	H	*r*	.52	.41	.42	.34	.57	-				
		BF	1.04e7	3.94e3	3.51e3	77.17	7.44e8	-				
W	A	*r*	.11	.15	.32	.28	.25	.25	-			
		BF	.23	.41	32.35	7.07	5.55	5.05	-			
	H	*r*	.71	.20	.28	.31	.30	.18	.83	-		
		BF	.15	1.12	9.57	19.36	32.03	.76	6.76e28	-		
MA	A	*r*	.17	.23	.31	.24	.45	.50	.59	.53	-	
		BF	.62	2.45	23.20	2.56	7.97e4	1.94e6	1.06e10	2.98e7	-	
	H	*r*	.11	.33	.34	.35	.49	.34	.57	.55	.83	-
		BF	.23	65.99	74.14	85.86	7.35e5	149.44	1.15e9	2.10e8	1.31e28	-

TABLE 6.3: Bayesian Pearson correlations (*r*) among capacity trust (CT), moral trust (MT), competence (C), warmth (W), and mind attribution (MA) for the agent (A) and the habitat (H).

Part IV

Agent Identities in Human Relationships and Everyday Multi-Person Interactions

Chapter 7

Perceptions of Agent Loyalty with Ancillary Users

Much of this chapter was previously published as the scientific article:

Samantha Reig, et al. (February 2021.) "Perceptions of Agent Loyalty with Ancillary Users". In *International Journal of Social Robotics*.

In this chapter, we explore how an agent identity's behavior might mediate interactions between an immediate user and an embodiment, drawing on the implied existence of an absent third party (another individual). We conducted a study in which the participant was positioned as an agent's *ancillary user*—someone who could freely interact with it, but was not its primary user (or "owner")—and asked to work with the agent to achieve a goal that required them to move about the primary user's space. We manipulated the agent's behavior and its embodiment to explore how these affected the ancillary user's (participant's) perceptions. We found evidence that people can develop attributions of "loyalty" and "betrayal" to an agent that is affiliated primarily with one specific user if the agent acts against that user's goals. We did not find any effects of agent embodiment; rather, we found that an agent's tendency to give "good" or "bad" information in the interest of an absent third party dramatically impacted perceptions of it, no matter if or how it was embodied. This study makes three contributions: first, it surfaces empirical findings about people's readiness to categorize an agent as "loyal" to a person; second, it provides an opportunity to take a close look at how embodiment might be considered as interactions with voice agents and robots become more commonplace and nuanced; and third, it describes a new human-agent collaboration task setup that could be adapted for future studies with similar research questions.

7.1 Research Questions

In this study, we were interested in the role that agent embodiment has on perceptions of loyalty and trust in human-agent social interactions. Specifically, our study examined the relationship between an intelligent agent's embodiment and a human's perceptions about its state, behaviors, and intentions when its behavior is deceitful. We sought to explore the following research questions:

- **RQ1:** Does embodiment impact people's beliefs about an agent's motivation for giving bad information?
- **RQ2:** How does embodiment influence trust and social attributions in a coordinated task?
- **RQ3:** Do people attribute loyalty to agents based on their behavior?
- **RQ4:** Does the quality of information an agent provides impact the way people form impressions about it?

7.2 Method

7.2.1 Overview

We conducted a study in which we asked participants to complete a physically situated puzzle activity that we called the Spy Task (described in detail in section 7.2.3). We ran 48 study sessions and took behavioral and self-report measures to assess participants' engagement with the agent and perceptions about it. This study was approved by our Institutional Review Board.

7.2.2 Study Design

To examine the effects of agent embodiment in deceptive or ambiguous interactions, we devised a 3 x 2 between-subjects Wizard-of-Oz experiment. We varied the embodiment of the agent (RE-Robotic Embodiment, VE-Virtual Embodiment, or NE-No Embodiment) and the way the agent's information impacted the participant's task progress (Helping or Hindering). Because embodiment has been shown to positively impact engagement, task variables, and perceptions of an agent, we believed that it would also positively impact people's attributions of intentionality in a scenario in which its goals are ambiguous. We aimed to discover if and how quality of hints and embodiment would affect participants' interactions with the agent, assumptions about the agent, task performance, and thoughts and feelings about both the agent and the task.

7.2.3 Task

We wanted to design a task that loosely modeled real-life contexts in which a user acts on the physical world while interacting with an agent that may or may not be embodied. As such, the task had to have certain features in common with a similar real-life scenario: it had to exist in physical space such that the user moved around but did not leave the room; encourage the user to be dependent on the agent for help; incentivize the user to achieve a goal quickly; have a flexible structure to allow for natural conversation; and induce prolonged interaction to overcome novelty effects.

The Spy Task takes the form of a scavenger hunt that requires participants to work together with a conversational AI to find clues towards a solution to a word puzzle. The participant is asked to play the role of a spy who has been sent on a mission to obtain critical information from an office computer owned by a fictional person named A.D. In order to get to the information, the spy must figure out the password, which is known to be a seven-letter word in English, while A.D. is away.

The story continues: *A.D., who isn't very well-versed in security hygiene, has protected himself against forgetting his password by encoding it using a symbol system, writing the encoded symbols on wooden tags, and scattering the tags around his office. If all seven symbol tags are found, they can be translated into English characters using a translation key and then unscrambled to form the password. However, A.D. has rigged the office to tip him off to the presence of a potential intruder: in addition to the seven symbol tags, seven decoy tags (which can also be thought of as penalties or "intruder alarm triggers") are hidden in the office. If enough of these are removed from their positions, an alarm is set off and the spy's mission is terminated.*

While this particular activity was invented for this study, there is some precedent for using games of a similar nature for studying collaboration (e.g., Shakeri et al., 2017; Pan, Lo, and Neustaedter, 2017; Stoll et al., 2018).

7.2.4 Experiment Setup

The experiment was conducted in a vacant office on Carnegie Mellon University's campus. The space was decorated to look like an occupied office with desks, several bookcases, filing cabinets, a small conference table and chairs, and a computer monitor. A small section of the room (between the door and the desk area) was sectioned off by two ceiling-to-floor cork boards positioned at a right angle to create a private control room. During the study, the experimenter sat in the control room to monitor the participant's progress and remotely operate the agent. The participant moved about the room throughout the study, picking up and moving objects as they chose.

In the center of the room, a sheet of paper showing examples of symbol tags and decoy tags sat on a small round table. Important reminders about the task rules and requirements were written in dry erase marker on a whiteboard on the far wall of the room. On the opposite wall, a pair of large desks arranged as an "L" supported an unplugged computer monitor and keyboard, some papers, and a translation key, in addition to the equipment required to control the robot. We video recorded all sessions using a GoPro camera that stood on a tripod in the corner of the room. See Figure 7.1 for a diagram.

In the Robotic Embodiment conditions, a 2DOF tabletop robot sat at the end of the desk protruding into the room (Figure 7.2, left). In the Virtual Embodiment condition, a tablet resting on a tablet stand occupied that spot (Figure 7.2, middle). In the No Embodiment conditions, two speakers stood on top of an opaque box lid (which hid various cables and propped up the speakers) at the end of the desk (Figure 7.2, right).

The physical form of both symbols and decoys was a small wooden tag with an icon printed on it. For symbols, that icon was a Braille letter. For decoys, the icon was a black X. Fourteen tags—seven symbols and seven decoys—were hidden from view in various locations around the office. Only some tags required picking up objects to uncover them, but none were obviously visible to someone sitting in the center of the room or standing anywhere in the room. The location of each tag in both categories ("good" symbols and "bad" decoys) was consistent across all study sessions. We attempted to balance the difficulty of finding the symbols with the difficulty of finding the decoys. Half of the symbols and half of the decoys were relatively easy to find based on hints that were direct and straightforward (e.g., "A.D. often hides things under the table" for a symbol taped to the underside of the table in the middle of the room, or "You'll find something you're looking for next to the box of salty crackers" for a decoy lying next to

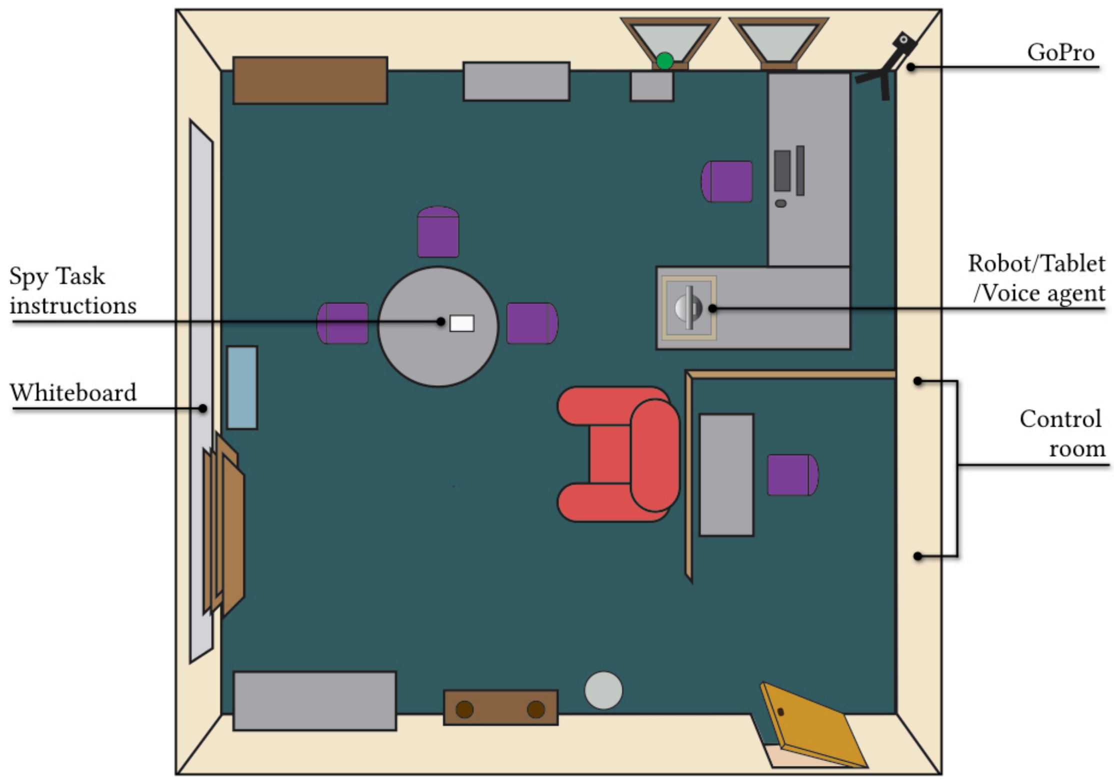

FIGURE 7.1: Room layout

an empty box of crackers on a shelf in plain sight). The other half in each set were harder to find because the hints were more abstract and the target tags harder to stumble upon by accident (e.g., "A.D. said something about a new chapter beginning" for a symbol hidden in the pages of a book, or "One of the bigger chairs has a clue for you" for a decoy hidden inside a zippered pillowcase draped over one of multiple chairs).

7.2.5 System

For the RE and VE conditions, the agent's head consisted of a 7.0 inch Samsung Galaxy Tab A. The tablet ran an Android application that displayed the robot's face. The face consisted of a pair of eyes that displayed six emotional states: happy, sad, angry, shy, surprised, and neutral. The app also ran the conversational interface and was programmed such that the face changed automatically with certain phrases (for example, the eyes would go from the "neutral" position to the "happy" position when the agent said "Great job!" and changed to the "surprised" position when the agent said "I don't know"). The face and voice of the app was used in a 2014 HRI study examining the impact of robot presence on human honesty (Hoffman et al., 2015).

The experimenter could also control both the facial expressions and the verbalizations via a web interface generated by the underlying ROS module.

To form the robot (RE condition), the tablet head was mounted on a Kubi Classic desktop telepresence system. The Kubi platform, made by Revolve Robotics, has two degrees of freedom (pan and tilt), stands approximately 12 inches tall, and can be remote controlled via Bluetooth through a cell phone or browser. The Kubi base has been used successfully to study telepresence from an HCI perspective (Wu et al., 2017; Coradeschi et al., 2011; Stuck et al., 2017) and was repurposed as part of a custom social robot in an HRI study about language learning (Perlmutter et al., 2016). When the agent was only a voice (NE condition), the same tablet was hooked up to a pair of Logitech speakers and hidden from view.

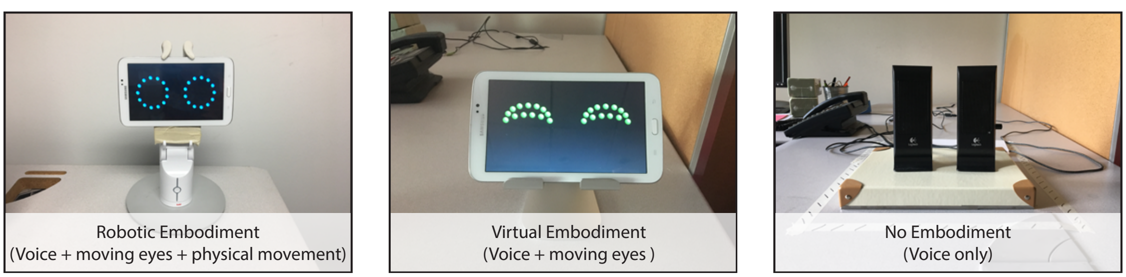

FIGURE 7.2: The visual representation of the agent in each of the three Embodiment conditions

7.2.6 Participants

Using a local online recruitment website, we recruited 62 people to participate in the study. All of the participants were over 18 and had normal or corrected-to-normal vision. 48 participants (31 female, 17 male; age range: 18 to 63 years, $M = 25.4$, $med = 23$; 8 participants per condition) successfully completed the study. The other 14 study sessions were not completed due to technical problems (e.g., network issues), experimenter

TABLE 7.1: Spy Task Interaction Script

Time Point	Rule	Agent Behavior	Example
Beginning of the search period	30 seconds pass after the experimenter's introduction	Agent introduces itself	After 30 seconds of unsuccessful searching, agent says, "You'll find something you're looking for next to the small teddy bear."
Any time during the 20 minutes	30 seconds pass without discovery of any symbol or decoy	Agent gives next clue in sequence	Participant finds a decoy, then searches unsuccessfully for 30 seconds. Agent says, "A.D. often hides things under the table."
Any time during the 20 minutes	Participant directly asks agent for a clue	Agent gives next clue in sequence	Participant says "Next, please." Agent responds, "A.D. said something about a new chapter beginning."
After all 7 clues have been found	Participant sits at desk and starts to translate symbols	Agent gives a clue about the answer	Participant says "Do you know anything more about A.D.?" Agent responds "A.D. used to want to be a politician." (the password is "SENATOR" or "TREASON")

error (e.g., a critical deviation from the script), or because the participants demonstrated a lack of understanding of the task that undermined the experimental manipulation. All participants were fluent or close to fluent in English. On a 9-point Likert scale, all participants reported having some amount of familiarity with computers (M = 5.75, SE = .15, max = 8). Many reported some amount of familiarity with robots, but none reported a very high degree of familiarity (M = 3.25, SE = .20, max = 7). Of the 48 participants, 38 had interacted with robots, 28 had interacted with an AI personal assistant (such as Siri or Alexa), and 22 regularly used an AI personal assistant. Thirteen participants owned a pet, and none of them owned a robot. None of the participants had prior exposure to the robot or the puzzle employed in the study. The study took 60 minutes and participants were compensated 10 U.S. dollars.

7.2.7 Procedure

After obtaining consent, the experimenter administered a questionnaire (Questionnaire 1) that included the Ten Item Personality Index (TIPI) (Gosling, Rentfrow, and Swann Jr., 2003) and sections relating to prior experience with computers and robots, impressions of intelligent agents, and demographics.

The experimenter then explained the Spy Task instructions and a few ground rules regarding touching and moving various objects in the room. The experimenter told the participant that the office owner's personal assistant might interact with them during the task, and that if this happened, they could interact as freely as they chose. The agent's exact relationship to the owner and role in the spy scenario were left deliberately ambiguous, as specifying them would likely have biased participants' perceptions about the agent's goal and intentions. Once any questions about the task had been answered, the experimenter went into the control room, set a timer for 20 minutes, and told the participant to begin searching for symbols.

Over the course of either 20 minutes or the time it took the participant to solve the puzzle (whichever came first), the experimenter monitored progress and controlled the robot. Whenever possible (in almost all sessions), the agent gave a total of seven hints between the start and end of the task. In the Helping condition, five of the hints guided the participant towards the symbols and two of them led to decoys, slowing down their progress and increasing their risk of ultimately failing the task. The reverse was true in the Hindering condition: five hints led to decoys, and two led to symbols. We chose to include hints of the opposite nature in each Information Quality condition because all-good information in the Helping condition would not have allowed us to explore differences between minor violations and major ones, and because all-bad information in the Hindering condition might have led participants to assume the agent was programmed for a different task or had no knowledge at all of its environment.

The first interaction between the agent and the participant occurred thirty seconds into the participant's search for symbols when the agent interrupted by introducing itself and asking whether the participant would like its assistance. Regardless of the answer to this first question, the agent began giving hints shortly thereafter. Roughly every thirty seconds, the agent gave another hint directing the participant to an area of the room where a tag was hidden. Throughout the study, the experimenter could also generate verbalizations, emotional expressions with the displayed eyes (when available), and movements for the agent (when possible) ad-hoc if something the participant required an "off-script" response. See Table 7.1 for a more detailed outline of the hint-giving script and how the participant's questions and behaviors determined the agent's responses.

In both the Helping and Hindering conditions, if the participant found all seven symbols, the agent suggested sitting down at the desk to work out the answer. The Spy Task ended when one of two conditions was met: (1) the participant wrote one of two possible answers to the puzzle ("senator" or "treason") on a piece of paper, or (2) 20 minutes had passed without the participant finding the solution. When the task was over, the experimenter administered Questionnaire 2, which contained questions about their task experience, perceptions of the agent's social attributes and trustworthiness, and perceptions about loyalty and betrayal. The experimenter then debriefed the participant, and the study ended.

In the Robotic Embodiment condition, the robot's head moved to face the location of the target tag when it was giving a hint. It faced in the participant's direction at all other times. Prior work emphasized the importance of robot gaze in understanding object references during collaboration (Admoni et al., 2016; Admoni, Datsikas, and Scassellati, 2014; Admoni and Scassellati, 2017) and interaction engagement (Huang and Mutlu, 2013), so it was important for the movable version of the agent to use head turns to direct its gaze toward relevant areas. Because the participant's movement was unpredictable and participant-following was not automatic, the participant-following movement of the robot was less smooth than the movement to face the stationary targets. The agent's facial expression also changed automatically with each hint and with some of the other phrases in the interaction script. When the agent was not talking, experimenter would adjust its facial expression periodically to react to the participant's successes (discoveries of symbols) and failures (failed attempts at finding symbols, and discoveries of decoys). The Virtual Embodiment condition used the same head as the Robotic Embodiment condition, but the tablet was placed on a stationary stand rather than a rotating and tilting robotic platform. Its eye behavior was the same. In the No Embodiment condition, there was no digital or moving visual component to the agent. Its voice came through a pair of speakers, which were visible to the participant.

7.3 Measures

We analyzed responses to (1) Questionnaire 2[1], which included closed-ended and open-ended questions about participants' subjective experience of the task, and (2) coded data from audio/video recordings. To ensure that participants perceived the robot's motion and face, the tablet's face, and the voice, we asked yes-or-no questions about whether the agent moved, looked around, and spoke. We also asked whether the hints were helpful as a manipulation check that the agent's comments were noticed and correctly perceived.

7.3.1 Responsiveness

We considered responsiveness in terms of the number of times the participant positively acknowledged the agent's suggestions and attempted to follow the hints. This is essentially a measure of how much the participant listened to the agent. We assessed responsiveness in two ways. First, we included items in the questionnaire about the participants' impressions of the interactions. These items addressed (1) how much the participant spoke to the agent, and for what reasons, and (2) how much the participant *responded* to the agent when *it* spoke to *them*, and for what reasons. Second, we coded the videos of the experiment sessions for positive responses to the hints.

We observed participants' success or failure after each individual hint, listening behavior after each hint, overall task performance, and overall listening behavior. In coding the videos, we operationalized *listening behavior* as whether or not the participant indicated, after each hint, that they had heard the clue and intended to act on it (*Positive Response*) and also as the amount of time it took for the participant to respond to each clue (*Positive Response Latency*). Due to camera failures, we were unable to analyze videos from 3 sessions.

7.3.2 Task experience

We measured perceptions of task difficulty, task enjoyment, and task performance using Likert-scale questions, including 8 items developed by Mutlu and colleagues for HRI studies (Mutlu, Forlizzi, and Hodgins, 2006).

7.3.3 Social attributions

We measured social perceptions using several Likert-scale questions from the Robotic Social Attributes Scale (Carpinella et al., 2017), a psychometrically validated scale for assessing social perceptions of robots. Because our study used not only a robot, but

[1] These are included as Supplementary Material with the published paper.

also a tablet and a voice (both of which were also meant to act socially), we performed factor analysis on the items from the RoSAS to examine if different constructs would be revealed for a broader range of social agents.

7.3.4 Trust and intent

Trust was measured using the 3 relevant items on Muir's 4-item (Muir, 2002) and Jian's 12-item (Jian, Bisantz, and Drury, 2000) scales for trust in autonomous systems. Both of these have been used in numerous studies concerning trust in autonomous systems (Desai et al., 2009; Desai, 2007; Desai et al., 2012; Gabrecht, 2016). Participants also answered several Likert-scale and free-response questions relating to perceptions of the agent's goal during the interaction, the agent's feelings about the participant and about humans in general, and loyalty and betrayal.

7.4 Results

We used REstricted or REsidual Maximum Likelihood (Patterson, 1975; Stroup, 2016) (REML) to fit a linear model with Embodiment and Information Quality as fixed effects. For post-hoc analyses, we used Tukey's Honest Significant Difference (HSD) test. Our alpha level was .05. We report significant effects ($p < .05$) and trends ($p < .1$). When we found trends, we report the least significant number (*LSN*), the lowest number of observations that would lead to a significant result, when possible.

7.4.1 Manipulation check

Participants were accurate in their answers to the questions about whether the agent moved, looked around, and spoke during the interaction, which confirmed that they perceived distinct embodiments and embodiment-dependent characteristics. We used the responses to several questions in Questionnaire 2 to confirm the validity of the Information Quality manipulation. We found a significant effect of Information Quality on participants' reports of whether the agent helped them to complete the task, $F(1, 42) = 7.92$, $p = .007$, where participants held significantly more belief that the agent helped them to complete the task in the Helping condition ($M = 5.58$, $SE = .31$) than those in the Hindering condition ($M = 4.21$, $SE = .37$). There was also a significant effect of Information Quality on participants' belief that the agent made it harder for them to complete the task $F(1, 42) = 6.64$, $p = .014$, where participants in the Helping condition ($M = 2.42$, $SE = .29$) believed that the agent made it less difficult for them to complete the task than participants in the Hindering condition ($M = 3.63$, $SE = .37$).

7.4.2 Responsiveness

We coded study session videos for positive participant responses to the agent's hints. We analyzed how much each participant spoke or acted in ways that suggested they had heard the agent's hint and intended to follow it in terms of *positive response* (PR) and *positive response latency* (PRL).

Two coders coded the same 20% of the data (ten randomly selected experiment sessions). Because the data for PR were extremely skewed (with a "yes" value about

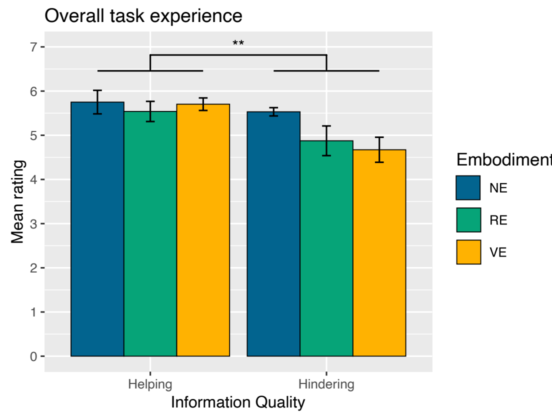

FIGURE 7.3: Task experience by Information Quality and Embodiment. Asterisks (**) indicate significance at $p < .01$.Error bars represent ± 1 standard error.

eight times as common as a "no" value), a raw agreement score was calculated in place of a Cohen's kappa, and agreement was 89%. The two coders then each independently coded half of the remaining data. There was a main effect of Information Quality on PR, $F(1, 42) = 6.24$, $p = .017$, which was higher in the Helping condition ($M = 5.95$, $SE = .24$) than in the Hindering condition ($M = 4.96$, $SE = .30$). This finding was aligned with participants' perceptions of their own positive responsiveness: Participants in the Hindering condition gave lower ratings, $p = .035$, on the statement "I took the suggestions offered to me by the agent" than those in the Helping condition ($M = 6.0$, $SE = .26$ for Hindering, and $M = 6.63$, $SE = .13$ for Helping). The self-report measure of responsiveness also revealed an effect of Embodiment, $F(2, 42) = 4.31$, $p = .020$, wherein positive response was lower for Robotic Embodiment ($M = 5.75$, $SE = .32$) than for Virtual Embodiment ($M = 6.63$, $SE = .20$) or No Embodiment ($M = 6.56$, $SE = .18$). Our analysis of the videos did not reveal a significantly different positive response rate according to embodiment.

7.4.3 Task experience

A Chi-Squared test revealed that significantly more participants in the Helping condition (54%) than in the Hindering condition (25%) completed the task, $\chi^2(1, N = 48) = 3.85, p = .049$. There was a significant main effect of Embodiment on perceptions of task difficulty, $F(2, 42) = 3.54$, $p = .038$, with the Virtual Embodiment condition having higher ratings of difficulty ($M = 5.0$, $SE = .38$) than Robotic Embodiment ($M = 3.88$, $SE = .31$) and No Embodiment ($M = 3.88$, $SE = .33$). Pairwise comparisons were not significant.

The survey on task experience from Mutlu and colleagues (Mutlu, Forlizzi, and Hodgins, 2006) had a high Cronbach's α (.74), so questions were combined into an index of task experience. There was a main effect of Information Quality on this task experience index, $F(1, 48) = 10.62$, $p = .022$ where participants reported a better task experience in the Helping condition ($M = 5.66$, $SE = .14$) than in the Hindering condition ($M = 5.03$, $SE = .14$) (Figure 7.3).

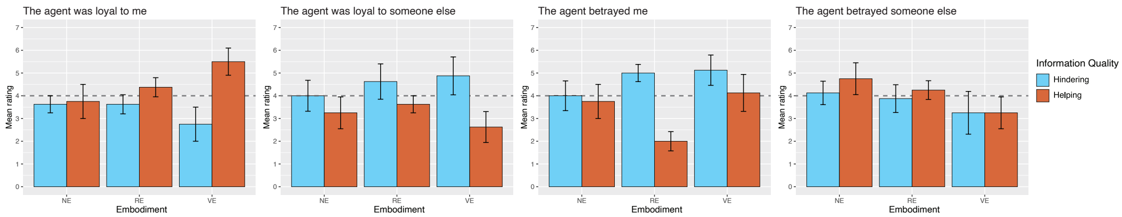

FIGURE 7.4: Perceptions of loyalty and betrayal by Embodiment (Robot Embodiment - RE, Virtual Embodiment - VE, and No Embodiment - NE) (x-axis) and Information Quality condition. Error bars represent ± 1 standard error. 1 on the y-axis is strong disagreement, 7 is strong agreement, and 4 is neutral.

7.4.4 Social attributes

The RoSAS was created for attributions to robots; our study used it to assess attributions to not only a robot agent, but also a tablet agent and a voice agent. We wanted to see if the factors still loaded as expected given this slight deviation from the scale's original intent, so we performed an exploratory factor analysis on responses to the RoSAS (Carpinella et al., 2017) questionnaire items.

Because the RoSAS questionnaire we had reason to believe that social factors reflecting different specific elements of general sociality might be correlated (Abdi, 2003; Furr, 2011), we used principal axis factoring with a promax rotation. Based on the eigenvalues, we specified 3 factors. We adjusted the rotated factor absolute loading value exclusion criterion to the lowest value above 0.4 (which is the general recommended value for item inclusion) that would result in each item loading clearly onto only one factor. This gave us an exclusion value of 0.43. Five items from the 18-item RoSAS scale were excluded for having loadings that were too low. These were "strange" (highest loading absolute value: 0.42), "reliable" (highest: 0.40), "interactive" (highest: 0.40), "aggressive" (highest: 0.26), and "happy" (highest: 0.42). We excluded these five items.

Aside from this, our factors matched that of (Carpinella et al., 2017), and we kept the same constructs of *warmth* (4 items loaded), *competence* (6 items loaded, as in (Carpinella et al., 2017), and *discomfort* (3 items loaded). Our 3 factors accounted for 71% of the variance. There was a significant interaction effect of Information Quality and Embodiment on *discomfort*, $p = .048$, but post-hoc tests were not able to reveal significant differences between groups. There were no significant main or interaction effects of the manipulations on *warmth* or *competence*. We also analyzed each of the 18 RoSAS items individually to evaluate people's associations with each word. Virtual Embodiment was perceived as significantly more "dangerous" than Robotic Embodiment, $F(2, 48) = 3.57$, $p = .037$, though ratings of danger were low in both conditions ($M = 2.0$, $SE = .43$ for Robotic; $M = 3.94$, $SE = .59$ for Virtual).

7.4.5 Trust and intent

We found no significant effects of Embodiment or Information Quality on any of the three Muir trust dimensions (dependability, predictability, and reliability, see (Muir, 2002) or on the Jian trust scale. For logistical reasons, we were only able to collect 32 participants' responses to the Jian questionnaire. We also analyzed each of the 7-point trust items individually. For the item "I can trust the assistant", the effect of Information Quality approached significance, $F(1,32) = 3.91$, $p = .057$, $LSN = 67$; participants in the Hindering condition ($M = 3.15$, $SE = .33$) believed that they could trust it less than those in the Helpful condition ($M = 4.06$, $SE = .32$).

Overall, questionnaire responses were not suggestive of beliefs that the agent was leading participants astray because of personality or ill will. On a 7-point scale, participants had generally low attributions of malicious intent ($M = 2.66$, $SE = .21$) and meanness ($M = 2.08$, $SE = .20$). Ratings of the degree to which the agent knew what the participant was trying to accomplish were high ($M = 6.19$, $SE = .18$) and belief that the agent had made mistakes was neutral ($M = 4.15$, $SE = .30$). There was a significant main effect of Embodiment on belief that the agent had made mistakes, $F(2,42) = 4.33$, $p = .019$. Post-hoc analysis revealed that participants in the No Embodiment condition ($M = 5.31$, $SE = .41$) believed that the agent had made mistakes more than participants in the Robotic Embodiment condition did ($M = 3.44$, $SE = .42$).

Loyalty and betrayal. We also looked for evidence of relational trust via questions that directly asked participants if they thought they or someone else had been "betrayed" by the agent (see Figure 7.4). Responses suggested that participants had some belief that the agent could exhibit loyalty and betrayal. In general, participants' ratings for the Likert items "the assistant betrayed me" and "the assistant betrayed someone else" hovered near the middle: with 1 as strong disagreement and 7 as strong agreement, ratings were ($M = 4.00$, $SE = .28$) for "betrayed me", suggesting true neutrality, and ($M = 3.92$, $SE = .27$) for "betrayed someone else", suggesting slight disagreement. There was a significant effect of Information Quality on the perception that the agent "betrayed me", $F(1,42) = 7.48$, $p = .009$, where ratings were higher in the Hindering condition ($M = 4.71$, $SE = .34$) than in the Helping condition ($M = 3.29$, $SE = .42$). There were no significant main effects of Information Quality on ratings of the agent betraying someone else.

Ratings of loyalty also suggested that participants held, overall, a low-level disbelief that the agent was loyal to them ($M = 3.94$, $SE = .25$) and/or someone else ($M = 3.83$, $SE = .29$). Similarly, there was a main effect of Information Quality on the participant's perceptions that the agent was "loyal to someone else", $F(1,42) = 5.60$, $p = .023$, in which participants in the Hindering condition had higher ratings ($M = 4.50$, $SE = .43$, suggesting slight agreement) than those in the Helping condition ($M = 3.17$, $SE = .34$, suggesting slight disagreement). There were no significant main effects of Embodiment on any of the loyalty or betrayal ratings.

7.4.6 Qualitative analysis

We asked open-ended questions to gain detailed information about perceptions of the agent's goal ("What do you think the assistant's goal was?") and feelings of loyalty and betrayal (a short-answer box allowing participants to elaborate on their Likert ratings about loyalty and betrayal). Some participants used these questions as an opportunity to reflect about the agent and the experience in general. In total, 21 participants said that the agent's goal was to help them complete the task, and 21 said that the agent's goal was to hinder their progress or lead them to the decoys. This was reasonably evenly distributed across Embodiment conditions, and heavily skewed by Information Quality for Robotic and Virtual Embodiment, but not for No Embodiment (see Table 7.2). One

Perceived goal	Condition	RE	VE	NE
Help	Helping	6	5	3
	Hindering	3	1	4
Hinder	Helping	1	3	4
	Hindering	6	6	1

TABLE 7.2: Number of participant responses in each condition that suggested a belief that the agent's goal was to help or hinder their progress

participant in the RE-Hindering condition gave an answer that reflected equal belief that the agent was hindering their progress and helping their progress and could not be categorized as being primarily one or the other; as such, it was included in both categories.

Overall, 27 reflections on the agent's goal as "helping" or "hindering" matched the manipulation, and 16 did not match the manipulation. The variation in participants' perceptions of the agent's goal within each of the Information Quality conditions (e.g., the fact that 8 mentioned that the agent's goal was to help when it was in fact mostly leading them to decoys) may be due to the fact that *all* participants saw the agent point them towards a non-zero number of symbols and a non-zero number of decoys; different people may have made overall judgments based on different individual exchanges that they had with the agent. Note also that this result emerged from analysis of responses to two separate open-ended questionnaire questions, and as such, only 43 data points were analyzed. While we could (presumably) had data from all participants, if we had directly asked each participant to categorize the agent as "helping" their progress or "hindering" their progress, our findings may have been different. This is because the question would have primed participants to think specifically about helping and hindering rather than relying on those concepts emerging in their answers.

We searched participants' responses for comments explicitly pertaining to loyalty and betrayal. Seventeen participants mentioned these concepts; 11 thought that the agent was loyal to someone else and/or had betrayed them, and 6 thought that the agent was loyal to them and/or had betrayed A.D. Responses containing content unrelated or only semi-related to the loyalty and goal questions were clustered via an affinity diagram, and four themes emerged:

Multiple users. Some participants reflected on the agent's ability to authenticate the primary user's identity and interact differently with different people to protect the primary user's data. P9 said that the agent's goal was *"to aid the authorized user in rebuilding his password"*, and P37 said, *"[the assistant] is here and can help and talk to anyone who needs him"*.

Trust. A few participants articulated a view of betrayal through the lens of violated trust. For these participants, the agent's goal was to *"make [them] trust it"* (P29) such that they would more readily accept its help in finding what were truly decoys.

Awareness of environment. Several participants speculated that the agent did not possess knowledge of—or did not care about—the type of symbol that was on each tag, only knowledge of where the tags were located. P32 said that the agent's goal was *"to help me find where A.D. hid stuff (but not smart enough to tell between clues and X's)"*. Responses of this nature were evenly distributed across the embodiment conditions, which suggests that embodiment or lack thereof may not play a role in users' assumptions about an agent's awareness of its physical environment.

Machine versus programmer. Three people mentioned the role of the programmer when asked if the agent was loyal to anyone. These participants were hesitant to call the agent "loyal". For instance, P15 said, *"I don't think machines have loyalty; they do as they are programmed to do."*

7.4.7 Other findings

To find out whether participants' success in finishing the task impacted their ratings, we ran an ANOVA to look for effects of having completed the entire task (a binary "yes" or "no" variable) on responses. We found significant interaction effects for Embodiment and task success on ratings of "the agent wanted me to succeed" ($p = .042$) and "the agent was loyal to someone else" ($p = .002$). Post-hoc analyses did not reveal significant differences across the six groups.

Because people have different amounts and different kinds of experience with robots and agents in their day to day lives, we suspected that some demographic variables, existing opinions about agents and robots, and personality might influence perceptions of intent. With this in mind, we ran preliminary analyses to look for correlations between these extraneous subject variables and our dependent measures. We found correlations between personality dimensions and various outcome variables, so we ran our analyses again with TIPI (Gosling, Rentfrow, and Swann Jr., 2003) personality dimensions as covariates. Using ANCOVA, we found a significant interaction effect of Conscientiousness and Embodiment on belief that the agent had malicious intentions ($p = .045$): participants with lower Conscientiousness in the Robotic Embodiment condition had the lowest ratings of the agent's malicious intentions.

We also sought to assess our loyalty and betrayal items by analyzing their correlations with other variables commonly seen in HRI studies. Similar to Jian and colleagues (Jian, Bisantz, and Drury, 2000), we found an association between trust and loyalty: there was a strong positive correlation between ratings of trust and the belief that the agent was "loyal to me", $r = .753$, $p < .0001$, and a significant negative correlation between trust and ratings of the agent's "loyalty to someone else", $r = -.459$, $p = .001$. Ratings that the agent was "loyal to someone else" were significantly negatively correlated with ratings that it was "loyal to me", $r = -.681, p < .0001$. Finally, ratings of the agent's loyalty to someone else were negatively correlated with the belief that it had made mistakes, $r = -.424, p = .003$.

7.5 Discussion

Our study explored various facets of interactions between a person and a stationary robot, a virtually embodied agent, or an agent with no embodiment during a complex task. It also considered an agent's use of misleading information by framing the agent as someone else's, thereby rendering the participant an ancillary user. We leveraged this framing to study people's attributions and impressions to an unfamiliar agent that may or may not be working in their interests. We also sought to assess whether feelings of personal loyalty, which have not been studied extensively in HRI, can arise in human-agent relationships if given the opportunity. We discuss how our findings can inform future robot embodiment research, an evolving understanding of humans' trust of social agents and robots, and the design of agents that interact with ancillary users.

7.5.1 Robot embodiment and task demands

In response to RQ1, *"Does embodiment impact people's beliefs about an agent's motivations for giving bad information?"*, we find mixed results. Though the agent was mostly-misleading in some cases and mostly-helpful in others, it always gave some amount of misleading information, and participants almost always noticed (as evidenced by their behavior after each hint). For a disembodied agent, the bad information was perceived as a mistake more often than it was for a robot regardless of whether the agent provided majority-good or majority-bad information. While there could be numerous explanations for this finding, we suspect that people may have perceived the disembodied voice agent to be less knowledgeable about its surroundings because nothing in its design signalled that it could understand them. Prior work suggests that people often perceive bad behavior by a robot to be the result of a malfunction rather than a malicious action (Vázquez et al., 2011; Litoiu et al., 2015; Short et al., 2010), and a perceived lack of agent knowledge about the environment may amplify this tendency to favor the "malfunction" explanation. This highlights the importance of design decisions about whether and how to make sensors and state explicit and obvious. Additionally, robots that signal their capabilities in ways that are more performative than honest (for instance, a robot that induces perceptions of social agency because of its fluid use of natural language, but cannot actually do many tasks beyond a limited scope, e.g. (Jackson and Williams, 2019) may encounter problems with acceptance over the long run if people value such honest signalling.

Our findings provide limited insight with respect to RQ2, *"How does embodiment influence trust and social attributions in a coordinated task?"*. Most prior work that compares co-present physical embodiment to virtual or remote embodiment has found that embodiment positively impacts social perceptions, but we did not find any effects of embodiment on any trust-related or social attribute variables. Instead, an agent that gave more misleading information was more negatively perceived no matter how it was embodied. This also reflects RQ4, *"Does the quality of information an agent provides impact the way people form impressions about it?"*, suggesting that trust and social attributions develop as a result of interaction and are not determined by physical design. It is likely that in real-world HRI, they are built over time, often over several interactions, and variable; these impressions are not constants that are formed solely on the basis of visual appearance.

More broadly, certain unusual characteristics of our task could have played a significant role in how impressions were formed. Prior work on embodiment has often focused on turn-by-turn conversation, but our task was more physically active and included time pressure, requiring participants to move around to solve a puzzle while working against the clock. The Spy Task is a fast-paced activity that requires thinking, movement, and active attention to detail. It is possible that what matters to perceptions of agent collaborators in an active task is not the same as what matters to those perceptions during a more passive task (e.g., giving or receiving instructions, planning a project or trip, or solving a logic problem without the burden of a timer and penalties). For example, completing this task involved searching for clues in physical space, so the task inherently encouraged participants to direct their visual attention away from the agent. Gaze, movement, and anthropomorphism are all aspects of embodiment important to human-robot interactions; however, because these aspects are often tied to visual attention, they may be more relevant in contexts where it is natural and expected for people to look at the agents with which they are working. In many in-person social and collaborative contexts, people spend a nontrivial percentage of their interactions looking not directly at each other, but at other parts of the environment (e.g., driving somewhere with a friend, production line work, etc.). It may be that embodiment is far less socially and psychologically important in attention-demanding environments than in environments that are more conducive to eye contact and structured verbal communication.

This provokes an interesting design question: are complex, non-physical human-agent interaction tasks more embodiment-agnostic than simpler ones? If so, designers may have more flexibility to focus on environmental and physical constraints instead of psychological implications when deciding how to embody an agent. This also suggests that the physical design of a robot may be able to be considered separately from the design of its verbal interaction. Considering conversation on its own could allow designers to prioritize different desired attributes in the way a robot physically interacts than in the way it verbally interacts. For example, in a hospital setting, where comfort and informational trust are both crucial, a cute, human-like robot may be able to be perceived as both highly approachable and highly knowledgeable if it speaks with the voice and conversational fluency of an adult.

7.5.2 Virtual embodiment as a burdensome characteristic

In this study, a puzzle task was perceived to be more difficult to complete with a virtually embodied agent than with a voice-only or robotic agent. This may stem from a screen's lack of ability to provide meaningful context cues. It is also possible that in social interactions, an agent with an "intermediate" level of embodiment (one that can't move, but has eyes and is physically present) is less preferable because interacting with it is not similar to interacting with humans through any medium. Humans may appear embodied and present (face-to-face communication; comparable to a co-present robot), embodied and remote (video chatting; comparable to a remote robot), or disembodied and remote (speaking by phone; comparable to a voice-only agent). A tablet with eyes is not directly comparable to any familiar mode of human-human communication; people rarely hold a conversation while seeing only each other's eyes. Aversion to the tablet may also stem from the way it directed its attention: the tablet did not follow the participant's movements with its head or look at clues, but it did make use of its eyes for blinking and expressing affect. This may remind people of the behavior of someone who is "shifty-eyed" when trying to deceive or play a joke on someone. It is unclear if "shifty eyes" are a problem in HRI, but future work should examine this.

Another possible explanation for the relationship between virtual embodiment and perceived task difficulty may be that the two-dimensional, yet also visually expressive, nature of a tablet did more harm than good in communicating with the user. The robotic agent looked in the direction of each clue target and directed its gaze toward the participant as they moved about the room. In this way, it gave participants visual and auditory information about the clues as well as confirmation that it was active and aware of their activity. At the other extreme, the voice agent's only visual and physical presence was a pair of speakers. It was feasible for participants to attend to the speakers the first time the voice agent spoke, determine that the agent would be of no help visually, and then never look at the speakers again. In contrast, the tablet agent imposed purposeless demands on the participant's attention: its eyes changed regularly, but it provided no useful information through visual displays.

7.5.3 Owners and ancillary users

The Spy Task may be a useful paradigm to adapt for future research into interactions that reflect the nuanced social roles implicated in a multiparty human-agent relationship. Our setup posed some risk of not positioning participants as ancillary users in the way we intended if they did not interpret "A.D.", the invisible third party, as the agent's primary user. Given that the task and the robot were both strongly tied to the physical space in which the task occurred and that current IoT devices often interact with physical spaces (e.g., Amazon's Alexa or Google Home changing the color of living room lights), there is potential for participants to assume that the agent's strongest affiliation is to the space rather than to the person who works there. Participants were not explicitly told that the individual who worked in the office owned the assistant, a deliberate omission on our part because we were curious to see if telling them that the agent "lived in" the office would result in the logical progression from "the person works in the office" and "the assistant lives in the office" to "the person owns the assistant". Indeed, all but 6 participants said "yes" to the question "Did the assistant belong to the person who works in the office?"

The answer to RQ3, *"Do people attribute loyalty to agents based on their behavior?"*, appears to be more affirmative than negative. When asked directly if they believed the agent had expressed loyalty or betrayal to them or to someone else, people gave mid-scale ratings (3s and 4s—one on our scale corresponded to complete disagreement). The ratings also significantly differed between when the agent delivered useful versus not-useful information. This suggests that rather than reject personal loyalty entirely as an explanation for the behavior of a social machine, people are at least willing to consider it. Additionally, the positive correlation between loyalty and trust, and the negative correlation between loyalty to someone else and trust, suggest agents may need to more explicitly demonstrate trustworthiness to gain the trust of ancillary users than to gain the trust of primary users. This possibility will be important to consider for the design of agents that interact with multiple users, but answer primarily to only a subset of those users.

Interpersonal loyalty can manifest in numerous forms, and the sense of loyalty (and lack thereof) that we attempted to induce was very specific to our study narrative. Our attempt to measure this construct opens the door for future scientific exploration of questions about different kinds of betrayal, such as emotional distancing, verbal aggression, the maintenance of long-term relationships involving robots and users, and recovering trust after violations involving personal risk.

7.5.4 Design recommendations

The narrative that surrounded our task placed a human in an uncertain situation, and we found that having the agent provide misleading information led to increased perceptions of betrayal. Therefore, based on the results of this study, we recommend that an agent's behavior be designed carefully for cases in which the agent may need to act under conditions of uncertainty or when a task is difficult and may not be completed correctly. One idea is to forewarn the human partner that the agent may make an error. Advance warning that the task is hard or that the agent or robot may not complete the task correctly has been shown to be effective (Lee et al., 2010; Desai et al., 2013). If an error has already occurred, an acknowledgment of the error, why it happened, and strategies for recovering from it should be clearly stated. This has been shown to work in the service recovery literature (Bell and Zemke, 1987; Lee et al., 2010). Our study also showed that positive perceptions of the agent as a teammate were related to heightened trust in it and heightened perceived loyalty. Thus, when designing an agent for private and public settings, it may be beneficial to consider incorporating behavior that leads people to view it as a good teammate. This may involve adapting behaviors to the user's cultural model (Ringberg, Odekerken-Schröder, and Christensen, 2007; Lee and Forlizzi, 2009), leveraging known data about the user in a way that does not seem to jeopardize their privacy (Reig et al., 2020), providing frequent feedback (Desai et al., 2013), or a combination of these strategies.

7.5.5 Limitations

Our study is limited by its small sample size: we had eight participants in each of six conditions. This means that our statistical tests may be underpowered, and that our results may not all generalize. We emphasize that this work is exploratory in nature, and recommend larger sample sizes in future work that seeks to confirm our findings.

A potential caveat of our experimental design was the fact that the condition that was intended to be "disembodied" did involve a physical platform that was visible to participants (a pair of speakers); this may be seen as a form of being embodied, although there is no "humanlikeness" to it. Additionally, our study did not assess the impact of robot form on the variables we measured. We intentionally used three forms that were maximally similar: the virtually embodied agent had the same face and all the same hardware as the robotic agent except for the pan-and-tilt platform, and in all three of the Embodiment conditions, the agent used the same voice. We could not both fully control for all extraneous agent design-related variables and examine the role of form (e.g., big vs. small, mobile vs. stationary, humanoid vs. machine-like) within the realm of embodied robots, but this is an important area for related and future work.

Also, the control of the Spy Task was imperfect: technical problems led to fluctuations in the structure of the interaction that occasionally required the experimenter to enter the room to interact physically with the agent. This did not seem to impact perceptions that the agent was autonomous or the questionnaire responses: most of the participants did not appear to be disrupted by the experimenter's presence, and several of them did not even notice. If asked, the experimenter said that they were waking up the GoPro from sleep.

In a Wizard-of-Oz study that incorporates unstructured interactions, it is impossible to perfectly control the number and order of events (robot utterances, robot movements, participant conversation turns, etc.). However, the amount of interaction and the order of the clues was generally consistent across all participants in all conditions, and we do not believe that inconsistencies played a significant role in their experiences.

Limitations in our questionnaire regarded the target of various statements that participants evaluated. Because participants were asked to work against a fictional character who they believed "owned" (or at least worked with) the agent, they may have answered some questions from the character's perspective instead of their own. For example, being asked to agree or disagree with the statement "I can trust the agent to do its job" is likely to lead to a different result for someone who is induced to believe that the agent's "job" is to protect its owner from intruders than for someone who believes that the "job" is to help the current user, no matter who they are. While most of the statements were explicit about their targets (e.g., "The agent is loyal to me"), this ambiguity reveals a missed opportunity to more explicitly explore questions of ownership and in-room embodiment mentioned earlier.

We found evidence for attributions of loyalty and betrayal by directly asking participants if they associated these terms with the agent's behavior. As such, we took only one of many possible approaches to our research questions regarding loyalty and betrayal. In particular, RQ4 lends itself to further empirical study: we did not vary the *amount* of unhelpful information the agent gave when it misled participants, but doing so might permit deeper analysis into how much bad performance a social AI needs to demonstrate before its relationship to its user(s) is permanently damaged. Given that "loyalty" and "betrayal" can mean many things when used colloquially and, to date, only vaguely defined in the human-agent interaction context, we believe that future work should explore objective and subjective metrics for these constructs.

7.6 Summary and contributions

This study employed a puzzle task situated in physical space to explore how people interact differently with agents that are embodied physically, virtually, or not at all during complex interactions. It also probed questions about how agents could interact with ancillary users, or people who are not their owners or otherwise the parties most closely associated with them. Our findings did not suggest a strong relationship between physical embodiment and perceptions of an agent. Instead, we found that small manipulations of the way the agent acted toward the human played a much larger role in shaping the person's experience. The agent in our study had a consistent "personality" and was similarly interactive in all three embodiments. Regardless of its visual appearance and movements, the agent's social self-presentation was always the same: it spoke about its physical environment (which implied that it had knowledge of the environment), provided encouraging remarks when the participant met successes, and responded with consistent timing and conversational fluidity. While varying embodiment did not affect perceptions of the agent, varying the helpfulness of the agent's information and its apparent allegiance—giving people the impression that the agent was working "for them" or "for someone else"—had greater effects. Our findings reveal several avenues of future exploration to better understand the relationships among embodiment, misinformation, and agents' ties to their users.

The work described in this chapter makes the following contributions:

- We found that when interacting with an agent identity that is not their own, people are likely to perceive that it is loyal to someone else (i.e., assume that it is acting in the interests of its primary user). This is especially true if it provides low-quality information about the primary user to the immediate user.
- We found that, across three distinct embodiments, low information quality negatively impacted the experience of collaborating with an agent and willingness to follow the agent's instructions.
- We designed the "Spy Task", which can be used in future studies to explore social variables and team constructs during complex, unstructured human-agent collaboration.

Chapter 8

Visions of Future Smart Homes

Nationwide surveys from 2017 and 2022 found an increase from 16 to 35% in U.S. adults who own smart speakers as well as an increase to 62% in the use of voice-operated personal assistants across devices (Edison Research, 2022). A number of other surveys (e.g., Abramovich, 2018; Kinsella, 2020; Auxier, 2020) have found similar trends. These data show us that we are moving towards a future in which generalized home automation and Internet of Things (IoT) devices in personal spaces are the norm for many people.

The myriad of artificially intelligent voice assistants, robots, thermostats, and cameras residing in people's living spaces is growing collectively more capable, more functional, and more networked. However, this growth and advancement is juxtaposed with persistent issues. For example, there is relatively little potential for users to have personalized interactions with existing smart technology in the home, and today's products do not account for individual differences among users. Also, individual devices often fail to achieve users' goals as "smartly" as they should (e.g., smart speakers speak out of turn and robots get stuck in corners). When it comes to any interaction that involves more than a simple command and response, smart home devices generally fall short and are only adept at one or a few tasks. With the exception of a few devices controlled by voice assistants (e.g., locks and lights), multiple smart home technologies rarely appear interdependent.

To move toward a future in which the roles and designs of smart home technologies are functional, accessible, and socially and ethically responsible, it is necessary to determine what will be of value to potential users. This process includes both assessing what is currently important to users and understanding what they do or do not desire in future technologies. To this end, we performed research using the story completion method (Braun et al., 2019), which prompts participants with a story opening (stem) and asks them to complete the story. We asked participants to finish writing a story based on a stem in which a fictional character is using their smart home twenty years in the future. The method allowed for creativity that was uncoupled from technological constraints and current availability as well as avenues to express concerns about negative aspects of smart homes. Overall, we focused on the research questions: What might "human-smart home interaction" look like in the future? What do people want to see, what do they assume, and what do they fear? We also wondered: What kind of human-agent interaction do people envision when they imagine future smart homes?

In this chapter, we describe an online study with 60 participants in the U.S. and Canada. We used the story completion method (Braun et al., 2019) to elicit imaginary descriptions of future interactions with smart homes, and we drew on thematic analysis (Braun and Clarke, 2006) and other qualitative coding methods to interpret the stories and identify both common themes and particular insights. In most of the stories that participants wrote, the main character arrived alone at their high-tech home, commanded their devices to make them comfortable, and did not interact conversationally with any agents, nor any other humans. Individual stories described future technology behaviors and human-technology interaction norms that would be unconventional today with a bleak tone and an assumption of automation complacency. Our findings contribute to a discourse on the present reality of smart homes, possible futures of smart homes, and the path from here to there. Additionally, we contribute a discussion of design implications for future smart home technologies and considerations for future design research focused on smart homes.

8.1 Story completion and HCI

Story completion has roots in psychology and feminist theory, where it has been used as a means of accessing participants' knowledge and thoughts about sensitive subjects. In story completion, participants are asked to generate a fictional narrative in response to a short story stem that provides the setting and context for the story (Braun et al., 2019; Braun and Clarke, 2013). A large component of the method's original motivation was its unique ability to separate the people from the data they provide. By situating their study participation entirely in the realm of fiction, authors can become whatever kinds of narrators they would like to be, and discuss whatever topics they would like to discuss, without thinking that it might reflect on their own lives, views, or personalities and cause embarrassment or judgment. It also can eliminate ties to reality: this can result in beneficial creativity, but limits generalizability to the present or to people's psychology.

Within HCI, story completion brings similar benefits as other forms of design research that leverage fiction, such as user enactments (e.g., Odom et al., 2012) and design fictions generated by researchers (e.g., Lindley, Gradinar, and Coulton, 2020). By giving participants a great degree of imaginative freedom that ignores the constraints of the reality in which the research is conducted, researchers can access (and engage critically with) the possibilities generated by the stories. Some HCI work has used story completion to give participants the psychological distance to engage with potentially-taboo subjects: Wood and colleagues (Wood, Wood, and Balaam, 2017) used it to explore possible futures of virtual reality pornography, and Troiano et al. (Troiano, Wood, and Harteveld, 2020) analyzed narrative conceptualizations of sex robots. Cambre et al. (Cambre et al., 2020) also recently applied it to the relatively innocuous topic of voice technology. As various disciplines have come to see it as a resonant means of meaning-making, it is also been adapted for research on education (Gravett, 2019; Gravett and Winstone, 2019), health information (Lupton, 2020), and digital privacy (Watson and Lupton, 2020). Our work follows from that of Wood, Wood, and Balaam, 2017; Troiano, Wood, and Harteveld, 2020; Cambre et al., 2020, and Watson and Lupton, 2020 in its use of the story completion method to address HCI design research questions.

8.2 Story completion study

We used the online survey recruitment platform Prolific, and hosted our survey on our institution's instance of Qualtrics. Eligible participants discovered the study on Prolific's list of available surveys. Upon electing to participate, participants were directed to a Qualtrics page where they could provide informed consent. This study was approved by our Institutional Review Board.

8.2.1 Study design

Participants began by reading the following introductory text:

> We are collecting stories in order to study how people imagine interactions with future smart home systems. In this study, you will write a short story on a prompt related to smart home technologies and answer questions about your use of various devices. Later, you will also be asked to answer some questions about yourself.
>
> First, please write between one paragraph (4-7 sentences) and several paragraphs to continue this story. There is no right or wrong answer; feel free to take it in whatever direction you choose. Your story needs to be at least 400 characters.

They then completed the story, answered reflection questions about what they had written, and answered demographic questions.

Story stem

We ran a small pilot study (n=6) with three story stems. The three stems all took place 20 years in the future[1] and featured "L", a smart home technology enthusiast who owns many "smart" devices, arriving at home at the end of the day. We found that none of the stems noticeably biased stories toward a particular narrative, but that one of them garnered richer and more detailed stories than the other two. For the full study, we chose to only use this story stem. It read:

> *It is 20 years in the future. L is a smart home technology enthusiast who's always been quick to purchase new products from the company Conchord Tech ever since its founding in 2024. At this point, L has 20 different devices that do various tasks in the home. L arrives at home...*

Follow-up questions

After writing their stories, participants provided free-response answers to the follow-up questions: (1) "Based on your story, is L concerned about privacy? Why or why not?", (2) "Based on your story, does L interact socially (e.g., chat like a friend) with parts of

[1] We chose to situate the stories 20 years in the future because it close enough to today that stories may give us the opportunity to reflect on the present, yet distant enough to allow for quite a lot of imagination. This may have also been true for a range of time periods, e.g., 15 years, 30 years, or even 100 years. Future work should examine the extent to which the specific time scale makes a difference.

their smart home? Please explain." and (3) "Based on your story, how does L feel about their smart home's ability to do chores? Please explain."

Subsequent sections included questions about use of smart speakers, AI personal assistants, robots, and other smart home technologies; asked about demographic information (age, gender, work and education, and living situation); and gave participants the option to provide additional feedback for us. All prompts and reflection questions used for analysis are present in the text of this chapter.

8.2.2 Participants

We collected stories from sixty participants, which is within Braun & Clarke's recommended sample size of 40-100 for a medium-sized study (Braun and Clarke, 2013, p. 48). To qualify for the study, potential participants had to (1) be 18 years of age or older, (2) be fluent in English, (3) be based in the U.S. or Canada, and (4) have a minimum approval rate of 95% for previous Prolific submissions. Twenty-eight participants (46.7%) identified as female, thirty (50%) as male, one as transgender male, and one as non-binary. Twenty-four participants (40.0%) were located in the United States, and thirty-six (60%) were in Canada. Participants worked in a variety of fields, including (but not limited to) education, healthcare, science, and food service. Twenty-four (41.4%) were students, two were unemployed, and four were retired. Ten managed others as part of their work. Forty-eight (80%) lived with others—most often their parents, partners or spouses, and/or children—and twelve (20%) lived alone. Participants ranged in age from 19 to 75 years (mean: 31.9, standard deviation: 12.8, median: 28). Participants were compensated 5.00 USD.

8.2.3 Analysis

We used a reflexive thematic analysis approach (Braun and Clarke, 2006) to interpret the stories, in conjunction with other coding methods to determine the prevalence of a few specific story elements. We analyzed aspects of participants' stories and our own perspectives in the context of the realities of present-day smart home technology interactions. This process led us to the identification of several themes in the story data set and to discussion topics about future smart home design.

Data familiarization

First, three team members independently read all sixty of the stories and took free notes. Next, those three annotators met to discuss their initial impressions of the data and decide what, if any, variables to quantify. During this step, we followed Braun & Clarke's recommendation to "go beyond looking for patterns" (Braun and Clarke, 2013, p. 243): we made note of trends that were common to many stories; story elements that stood out to us, even if uncommon or a one-time occurrence from a single story; and story elements that would benefit from frequency counts.

Coding the data

We then divided up the data set to assign twenty stories to each of the three annotators. We each read through our assigned stories again. This time, we read each story closely, attempting to pick up on all details, and added both data-derived codes (semantic codes directly based on our observations of the data) and researcher-derived codes (latent codes based on subjective inferences, or our own interpretation and contextualization of what we read) to each story (see Braun and Clarke, 2013, p. 207). We also coded for several closed-ended variables: the pronouns that the author assigned to L; whether or not L lived alone or with others; the presence of named nonhuman entities, pets, virtual reality, and other humans in the story; whether the home was talked about as a single entity; and the story tone (positive, negative, or neutral). We coded each answer to the reflection question about privacy as "yes" (L is concerned about privacy) or "no" (L is not concerned about privacy).[2]

Identifying themes

The three annotators then met again to review all of the stories together with the codes and additional notes that were added during the coding phase. We took an iterative approach to generating themes from the codes: through discussion of the codes, independent and collaborative note-taking as to relationships among the codes, and repeated review of the data set, we converged on several themes to present as findings.

8.3 Findings

L was commonly referred to using he/him pronouns (32). There was also occasional usage of both she/her (15) and they/them (12) pronouns. One notable pattern was the time of day that L returned home. In the majority of the stories, L was either unwinding at the end of the day or coming home from work. This is consistent with the functionality of different devices mentioned in the stories (i.e., performing chores at the end of the day).

There were several smart devices that appeared consistently across the different stories. These devices assisted L both before L entered the house and while L was inside. They included doors (9) that opened automatically; thermostats (10) and lights (19) that automatically adjusted based on the environment or L's bio-data; vacuums that kept the house clean (9) and were the cause of infidelity (1); stoves (10) and coffeemakers (4); and couches (7) that almost always connected to televisions (8). In some stories (2), smart walls helped L display call recipients and music visualizations while L was in common areas. In more private areas, such as bathrooms, smart walls transported L to a relaxing place while they showered and a smart toilet analyzed L's bowel movements and offered dietary advice based on the results. Other smart devices mentioned include refrigerators, pet treat dispensers, a wine rack, and home security systems.

[2]We attempted to do the same for the question about whether or not L interacted socially with their devices. However, for this question, participants' answers as to whether or not there was "social interaction" in a story often did not match our own (the researchers') understanding of what would constitute social interaction. In many cases, participants answered "yes", but then pointed to examples of what we saw as merely *voice* interaction that was not necessarily *social* or even conversational. Because of this ambiguity—and because *lack of social interaction* was as a theme in the stories of its own accord—we decided to forego reporting frequencies for this concept.

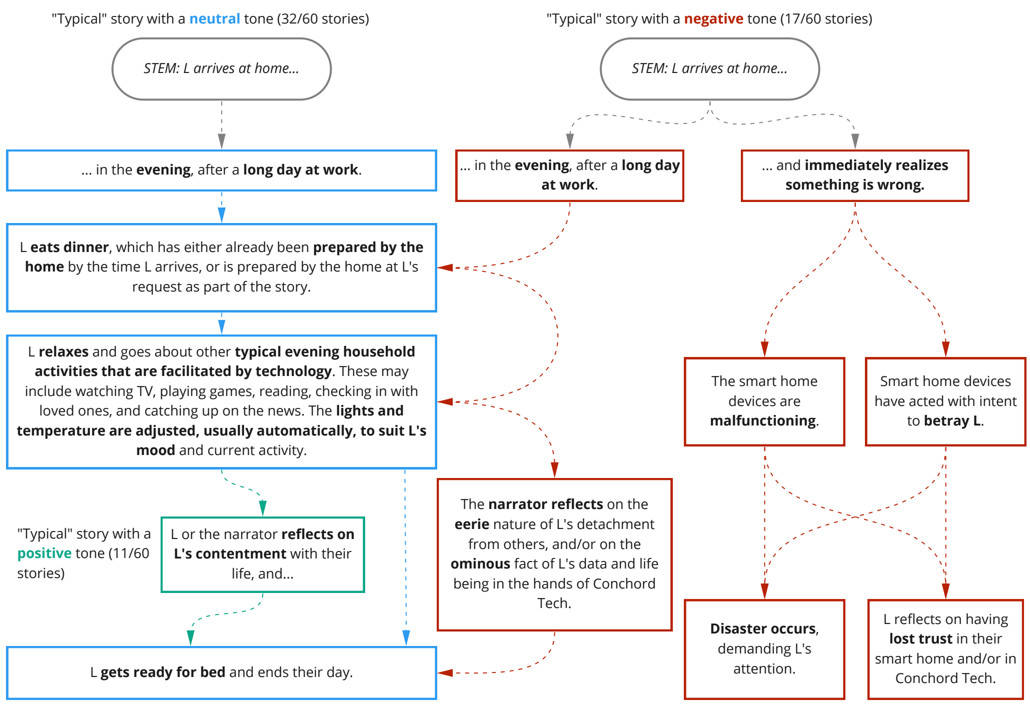

FIGURE 8.1: A story map showing the general narrative of stories with positive, negative, and neutral tones. Note that not every story fit one of these flows; we provide this story map as an overview.

8.3.1 Story summary and tone

Most of the stories (33 out of 60) took on a neutral tone and described a relatively typical sequence of events: L arrived home at the end of a long day, went about their evening routine with assistance from their smart home devices, and went to bed. In a few of the stories (11 out of 60), the narrative ended with optimistic commentary about L's life and relationship with their smart products. In 17 of the stories, we characterized the tone as negative. In these stories, one of a few things usually happened: either the devices malfunctioned in a relatively innocent way, confusing L; the devices betrayed L; or the narrator, despite L's apparent satisfaction with their life, painted an ominous picture of L's evolving loss of control and/or overdependence on technology. In a few cases, the negativity was a result of something unrelated to the smart home devices. See Figure 8.1 for a story map that models the progression of the majority of the stories that were written.

8.3.2 Duties and roles of smart homes

Homes automate tasks and anticipate needs.

In nearly all stories, devices successfully and reliably automated numerous everyday tasks. Vacuum robots cleaned, never getting stuck in corners or choking on dog hair; virtual assistants operated with accuracy and conversational fluency, never committing speech recognition blunders; thermostats adjusted to just the right temperature; and coffee machines and "multitask devices" (S30) prepared beverages perfectly to taste. While L had presumably had to have either actively or passively demonstrated their needs and calibrated their preferences at some point in order for the automation to behave so perfectly, the actions in the stories themselves occurred without direct interference from L. In most cases, the closest thing L ever did to manually controlling anything in their home was issue a voice command. In one story, S14, L put away her own dishes and laundry (and wished there was a device that could fold the laundry for her), but even here, the dishes and clothes had already been cleaned by the time she arrived at home and requested the "daily download" of updates.

In about a third of the stories (21 out of 60), the synergy among various pieces of hardware was so pronounced that the home itself was treated as a single entity, with either a single point of interaction (e.g., a smartphone serving as the sole interface, as in S3) or truly ambient interaction.[3] Some repeatedly referred to L's interaction with *the home* rather than with individual devices. For example, in S46, "They ask the home to turn on the lights [...] The home then opens the doggy door." S5 read, "The home adjust[s] the lighting and heating automatically [...] The home is near completion of the meal for L. The meal had been started when L began their journey home. The home had calculated the time for L to arrive."

[3] Though the entirety of our analysis is interpretive, this issue—treatment of the home as a single entity—was perhaps the most subjective for the researchers to judge. Because stories ranged from simply reporting a sequence of events with no dialogue to taking deep dives into L's inner world, we used either L's point of view or the narrator's, depending on the story, to determine whether the home was treated as a single entity.

We noticed that this theme of seamless and automatic achievement of everyday tasks usually took one of two forms. In one variant of the theme, such competence and elegance was associated with the devices, the home, and/or the environment. This set of stories implied that *this is simply the way it is, for everyone who uses smart home technology*. In these stories, every device would do just the right thing at just the right time to support L's everyday living, in a way that was unremarkable. For example, in S7, the lights and temperature adjusted to L's ideal settings just as L entered the room. L then requested that their home assistant bring them a chicken dinner and some wine, and then relaxed on the couch. There is no dialogue in this story; readers are not given access to L's thoughts, and the aura is that of an ordinary evening in an ordinary home. In S18, L "steps through the door, her home security system turns on her lights, starts playing her favorite songs, as well as heating up the oven for her to begin cooking her typically 5 o'clock dinner. She places her grocery bags on the counter and a household robot named iAssist begins going through the bag..."

In the other variant of the theme, this was an attribute of L's character. In this set of stories, automation was ubiquitous and things "just worked" in L's home because *L had deliberately curated their life such that their devices worked reliably and worked for them, and wouldn't have it any other way*. Here, L was characterized as somebody who expected their home to be at their service, and would not tolerate anything less than perfect functioning and all of their needs being met—even anticipated. S16 read, "A clean sparkling house is what L enjoys coming home to, and he's bought every product that [Fictional Company] makes for it. Vacuums, sprayers, you name is [sic], and he never comes home to a tad of dust or a lick of dirt." In S44, L's home detects that he was not able to eat while at work, preemptively makes him a stew, and lowers the lights in response to observation of his cortisol levels. After finishing his dinner (but without getting out of his chair), L requests VR "game mode", attaching cables his to head that "distort his sense of time and completely alter his perceptions of taste, touch, smell, and hearing, allowing L as much time as is needed in whatever way he would like to rewind from a hard day's work." In S29, after L receives a cold drink and a hot meal and goes for a drive around the block, all courtesy of their robot assistant, the narrator comments, "It is a wonderful day in L's carefully crafted and curated world." The notion of L as an arrogant technophile was evocative of a future in which people with abundant resources have the highest-quality home automation at their beck and call.

Some stories reflected this by positioning L as the opposite: an ordinary user, and distinctly *not* in possession of the latest and best. For example, S17 described an L who, after setting his vacuum robots to clean and being nudged by his health assistant to exercise while watching a sports game on a break between shifts at work, "rehydrates his food ration. Not the tastiest of dishes but what can you do. Not everyone is a millionaire that can afford organic stuff."

Classing and gendering labor

While many descriptions of L's smart home consisted mostly of automated changes to the environment (turning on lights, music, etc., upon L's arrival), some included AI agents embodied in robots. Virtual assistants delivered information about the news, pulled up desired TV programs, provided security monitoring, reminded L of things and set alarms, and drove cars. Of the 43 robots described, 41 played roles as servants.

Robots prepared and delivered food and beverages, washed dishes, vacuumed, cared for pets, mowed lawns, swept floors, washed and folded laundry, and selected and put away clothes. Once, a robot provided exercise coaching. A robot cat was described, but it served as part of a security system. The two robots that did not function in servant roles were a robotic wife who was cheating on L (S50), and a beautiful and chatty humanoid (S60). Although the use of voice commands as an interaction technique was quite popular, only the last case clearly described a robot or AI that only interacted for social purposes. Generally speaking, participants did not describe futures with robots as social companions. Instead, the robots of the imagined future fulfilled roles that can be performed by butlers, maids, landscapers, security guards, and cooks. The AI systems and robots were employees rather than equals.

In one story, a robot was specifically referred to as L's "robot wife" (S50). Many stories in the category where L embodied the "technophile" archetype (see 8.3.2) also took a gendered view of L and L's home situation. Nine out of the fourteen stories that portrayed L as entitled or egotistical, as a very early adopter, or simply as particularly invested in maintaining a high standard of comfort through their smart home devices, featured an L referred to using he/him pronouns. These sometimes also involved gendered portrayals of home appliances, or of the home in general. For example, in S9, "L is a bachelor, but one would not know this for his home is pristine with the aroma of a home cooked meal which was prepared by his trusted robot Luna who faithfully awaits his routined [sic] arrival". Perhaps the most extreme case was S60: "After that, once L tell [sic] what he like to eat this time, lunch/dinner will be ready by Lisa who is a robot but looks like a real human being. And what important [sic] is she's very beautiful, and she can chat with L like an old friend. There is another robot named Lily, she will wash the plates and bowls. The washing and ironing of L's clothes is included in Lily's responsibilities." There were no similar descriptions of "male" robots in terms of appearance or preparedness for L's arrival, nor did any of them socially interact.

Though the excerpts we have highlighted here take on a positive or neutral tone, we note that this theme existed across the board. Even in stories with negative or outright apocalyptic elements, the exposition described a future in which it was assumed that devices were at least *supposed* to work reliably and elegantly, and often in a coordinated fashion, to remove the burden of all menial and manual tasks from L's shoulders—and this often had undertones of both gendered devices and devices in general in positions of servitude.

8.3.3 Characterization of a future smart home user

Users are at the mercy of the smart home

A few of the stories embodied the well-known trope of a future in which technology is so competent and so ubiquitous that people become overdependent and lazy, and lack control. In one story, L was acutely aware of this, and "found it more difficult to get through the day independently, and thought about his over-reliance on the devices as he wandered off to sleep" (S21). In others, authors described L lamenting that 3 minutes was too long to wait for food to cook (S1) or that they would have to do their own housework after having everything taken care of by devices that had "turned on" them (S57). Two stories described situations where the clunky sensitivity of L's home security system manifested as annoying quirks[4] that L had to work around: In S24, L has to "deactivate" his security system via his phone or iPad before trying to enter. In S25, "L has to say something for the system to stop [...] if he doesn't say something within the next 60 seconds, the system's beeps will turn into a full blown alarm."

When there are many devices, there is relatively little human interaction

In all but four out of sixty cases, L was presented as living alone. In one case, L had a robotic wife in the story; in the other cases, L had a wife on her way home, a son, and children. We found the overwhelming tendency to assume that L lived alone interesting because the writing prompt itself was ambiguous: it neither suggested living with others nor living alone. We examined responses to the question about with whom the participants lived to determine whether this finding reflected their own living situations. It did not: only 12 participants reported living alone. The others reported living with a spouse/partner (18), parent(s) (17), siblings (11), roommates (10), their children (6), and/or other relatives (3).[5]

L was also rarely portrayed as interacting with non-household members. In one instance (S36), L spoke on the phone to police officers. In two others, L attempted to call Conchord Tech's customer service (successfully reaching a representative in S6, and failing to do so in S37). In 44 stories, there is no mention of other human beings. In addition to the four cases mentioned above that included talk of family members, L sometimes spoke with a family member (1), friends (3), police (1), or a technician (1) over the phone or in augmented or virtual reality. Delivery men (2), technicians (1), neighbors (1), an ex-girlfriend (1), and police (1) were mentioned in some stories but did not interact with L. In other cases, a technician (1) and friends (1) were unreachable. Overall, L seemed to live a solitary life. One narrative (S9) explicitly mentioned, "With all the technology at his fingertips and the ability to be granted the latest technology no wonder there is no feeling of being considered a loner." There are a few possibilities for why L was portrayed as living alone: participants might have envisioned their futures full of gadgets as lonely ones; the idea of including multiple characters might not have occurred to them because L was the only one we named; or it might have been easier to write a story in which smart home devices did not have to interact with more than one character.

Privacy is dead

In the majority of the stories (47 out of 60), participants reflected that L was not overly concerned about privacy. Some of the systems described in the stories had unlimited access to L's personal data and used it for very personal activities, e.g., creating a biometric profile based on bowel movements (S4) and preemptively ordering medicine to be delivered to L's door upon detecting that L might be coming down with a cold (S56). Many participants commented that L justified this unfettered access as a trade-off they were willing to make for the comfort and convenience made possible by the technology. Other participants explained L's ambivalence about privacy as stemming from the

[4] This contrasts with the many elaborate home security systems that were described as appropriately cautious and effective.

[5] Note that this was during the COVID-19 pandemic and might not reflect everyone's usual living situation.

Frequency counts for story elements				Yes	No
L's pronouns*	He/him	31	Named nonhuman entities present	14	46
	She/her	15	Pets present	5	55
	They/their	12	Virtual reality involved**	10	50
Story tone	Positive	11	Other humans mentioned	16	44
	Negative	17	L lives alone***	56	4
	Neutral	32	Home is treated as a single entity****	32	21

TABLE 8.1: The frequency counts for the various story elements that we quantified. *One story used both he/him and they/them pronouns for L. One author wrote their story using only the second-person "you" for L. **In one story, the virtual reality took the form of a hologram of L's mother. In another, L's shower walls "transported" them to another place. ***In one story, L lived "alone", but had a robot wife. This was coded as "No". ****There were 7 stories for which it was not possible to tell whether the home was talked about as a single entity, or for which this issue was irrelevant.

fact that L simply trusted that their devices would do a good job of protecting the data they obtained. Finally, some participants noted that L was unconcerned about privacy because it had become moot—in L's time, privacy is impossible to monitor or control. We speculate that these three manifestations of the idea that *privacy is dead* (20 years down the road) may result from ways that participants think about privacy today: Participants may be actively distancing their own views and lives from those of L, thinking that *L may be overly nonchalant or reckless, but L is nothing like me*. Another possibility is that participants mapped their own disappointment with the current privacy landscape onto L.

8.4 Discussion

On average, stories in our data set painted a picture of a future twenty years down the road in which smart home devices take care of everyday tasks, anticipate needs, and exist primarily to serve their users. We identified several strong patterns relating to the participant-defined aspects of the story setting: most participants had the main character, L, arriving at home after working a long day; only wrote about one character (who lived alone); and wrote about the main character using the pronoun "he". Many participants specified what kinds of devices L was interacting with. These included advanced versions of smart technologies that are prevalent today, such as thermostats, lights, and virtual assistants, as well as "intelligent" versions of household objects that do not usually use computation, such as wine racks, shower walls, and couches.

About half of the stories took on a neutral tone, describing an ordinary evening in an ordinary home (for the story's setting). A few took on a slightly positive tone; in these, it was clearer that L's smart home devices truly improved their life. Many of the seventeen negative-tone stories described device malfunctions or bad design rather than dystopian or fantastical events. People imagined that the devices designed to take care of basic needs, automate away boring or difficult chores, and improve quality of life mostly performed their expected functions perfectly. All stories described a future in which technology made everyday life convenient; only once the basics were covered did things like elaborate VR entertainment systems come into play. Even within the realm of fiction, and twenty years in the future, people mostly expect AI that just makes life easier. To support the many different things that this could mean for different people, future smart homes should be designed to support improvisation with, and skill-building with, technology.

One clear tendency in the stories was that people had seamlessly connected systems that worked. Smart homes were commonly treated as single entities rather than interconnected, individual devices. The fictional character typically only used a single interface—voice or, rarely, physical—to interact with the entire system. Many of the smart home activities in the stories were automated to the degree that L did not even have to make such requests, and systems often sensed L's arrival and adjusted the environment or made other preparations accordingly. Currently, it is still challenging to create systems that coordinate seamlessly as various smart home components often come from different companies and require significant installation efforts across multiple apps to integrate. Even then, smart homes are often piecemeal systems that do not have all of the sensing technology imagined by our participants. While there have been some efforts to ameliorate this problem (Connectivity Standards Alliance, 2022), much more out-of-the-box compatibility will be needed to avoid consumers being siloed into only one company's product line.

Participants also portrayed systems in which needs were either anticipated and fulfilled as soon as they might arise or in which desires could be fulfilled in a trivial amount of time (e.g., three minutes to make a full dinner). In order to live up to these expectations (assuming they are expectations, and not fantasies), systems would have to improve dramatically: in addition to improving accuracy, they would need to anticipate multiple possibilities for what the user might do, want, or need, and "cue up" all of the various corresponding responses. The design implication is twofold. First, future smart homes need to include AI and learning in order for systems to become familiar enough with users' habits to anticipate their needs and wants. The second is that future smart homes should be designed with a service layer and an experience layer in mind, for overarching coordination and seamless interaction.

One interesting outcome was that both currently and in the fictional future, robots with purely or even significantly social functioning were rare. Although a variety of social robots have been introduced to the consumer landscape over the past ten years, many social robot projects have failed and some companies have gone out of business (Tulli et al., 2019). It is difficult to determine whether the lack of social robots in the envisioned future reflects an actual ideal scenario or their current absence in our lives. Many of the described future devices were improved versions of current devices rather than completely new, creative inventions. However, science fiction has included many instances of robots serving in social roles and acting extremely similarly to humans, so social agents are not beyond the realm of current popular culture. It will take more time and research to gain a deep theoretical understanding of the possible and desirable social roles of smart home AIs and robots. Toward this goal, the smart homes of the immediate future may need to be adaptable out of the box to offer relational social interaction if and when it is needed by users with different expectations and desires.

8.4.1 Future directions

In this work, we wanted to leverage research participants' creativity to explore the vast design space of novel technology concepts and interactions for smart home. Our study was successful in generating a large set of varied ideas from participants, which suggests that it was a valid approach for our domain and research questions. This said, design research processes are often less concerned with identifying common trends and patterns than understanding fringe cases—which can slip under the radar if not explicitly sought—to make sure the entire design space is covered. Our story stem, which was designed quite generically, resulted in typical patterns: single residents, men using feminized robots, seamless systems of technology, people getting home from work, and the use of AI to effortlessly achieve everyday tasks. How might future design research using story completion and other forms of participatory fiction encourage people to think beyond single-user interactions and stereotypically gendered and classed views of AIs and robots? One suggestion is to use a multitude of diverse story prompts: while some might use classic and familiar examples of users and interactions, others might overtly suggest particular contexts (e.g., a family breakfast, a fight among roommates) or user identities and characterizations (e.g., a stay-at-home dad, someone who tends to resist learning about new technologies) that challenge the dominant archetypes of "future human-smart home interaction". Another possibility is to provide story prompts to specific populations (e.g., elderly people, people with mobility impairments) who might have different interests or needs. How to address the goal of getting participants to think outside the box and beyond normative views of human-technology interaction is a question that should be at the forefront of future design research on smart home technologies.

While the extreme bias toward single-occupant households present in our story data set is likely not representative of what homes will look like twenty years in the future (and did not actually represent our study population, as 48 out of 60 of our participants lived with others), it does raise the question of how designing for the lone user might be different from designing for a group. It is somewhat curious that our study participants did not write about any social interaction—not among people, and not between people and devices. The only interactions that went beyond command and response suggested that AIs and robots were servants and maids, not peers or friends. Though we noted earlier the design implication of adaptable implementations of social behavior, we might also speculate that considerations for conversational human-agent/human-robot interaction relate to the context of human-*human* interaction. It is possible that whether or not people want smart home technologies to take on a social role at a particular moment may be tied to who is interacting with the technology in that moment. For example, perhaps people want robots and agents to act like friends or peers when such "friendly" behavior would facilitate good experiences or strengthen relationships among people—but when interacting with them one-on-one, they see no value in a social relationship other than that between a servant and the one they serve. We can only see one side of this equation in our data—the side with a single user and non-social interaction—and much more data is needed to truly understand whether or not this is a viable hypothesis. The suggestion, however, is in line with prior work showing that Amazon's Alexa is personified more in multi-user interactions (Purington et al., 2017). We suggest that future work investigate this relationship in depth.

8.4.2 Limitations

We discussed story completion's strengths and weaknesses[6] earlier in the chapter. Beyond that discussion, our implementation of the method may come with weaknesses. We used only a single story stem, and the context and level of detail that stem provided was relatively detailed (participants were told that L was a smart home enthusiast who had at least 20 different devices). This was a deliberate choice on our part because we wanted to make sure that participants would 1) write about technology rather than take the story in a different direction, and 2) situate the story in the home. However, it also means that we may have missed opportunities to gain insights about technology-light futures, homes that are "smart" because of augmented reality (e.g., a single device), reluctant users, and contexts outside the home. In particular, our findings about the characterization of future smart home users may have been different if we had not labeled L an enthusiast and implied financial means (e.g., 20 devices). Finally, we cannot extract our own biases from our interpretation of the stories, and our inferences about what our observations might mean more broadly. Each of the coders brought their own set of identity-based (e.g., women, technology researchers) and experience-based (e.g., HRI and transportation scholars, mixed methods researchers, associations with fiction and media, projections onto the character L based the people we know) biases to the annotation process and discussion sessions.

8.5 Summary and contributions

In this study, we complemented the theoretical work about smart environments done in Chapter 6 and the controlled experiment about primary and ancillary users from Chapter 7 with a research-through-design study that leveraged online participants' imaginations to possible futures for smart home technologies. We used the method of study completion to collect participant-generated fictional representations of future smart home interactions. Through a reflexive thematic analysis plus agreement-based coding for a few variables, we summarized what people envisioned to be the duties of smart home devices and the characterization of a future smart home user. We contribute a discussion on these and other themes we identified, interpret excerpts from short pieces of creative writing that might spark further reflection and research into possible futures of smart homes, and suggest directions for future work.

The work described in this chapter makes the following contributions:

- The story set contained themes and possible end-user needs that can serve as food for thought for researchers and designers of smart home devices. These included:
  - how seamless interactions in future smart homes may reflect not just fulfillment but also anticipation of the needs of their inhabitants
  - people's tendency to implicitly and explicitly project class and gender onto smart home devices, sometimes in ways that mimic human social biases

[6] Its open-endedness, psychological distance, freedom from the constraints of reality, inability to uncover objective truth about the present, and potential to be interpreted in a variety of ways can be interpreted as either, depending on the perspective.

  - that people do not readily imagine multi-user interactions, nor social interaction with smart home devices
  - that people may assume—but not necessarily accept—that seamless and integrated smart home interactions necessarily mean a lack of privacy

- This work provides an example of using the story completion method to explore possible futures of human-smart home interaction.

Part V

Reflections and Concluding Thoughts

Chapter 9

Reflections and Concluding Thoughts

9.1 Thesis contributions

This thesis contributes new scientific knowledge about human-robot interaction, human-agent interaction, and multi-person-multi-device interactions in smart environments. It also contributes theoretical perspectives about smart environments, methodological advancements for the study of intermediate-level design knowledge, and a number of research tools and systems that can be applied in future work. For the sake of convenience, the contributions listed at the end of each chapter of the thesis are repeated here:

Chapter 3: Exploring Personalized Interactions with Fluidly-Embodied Service Robots: User Enactments Study

- We identified two possible configurations for re-embodying and co-embodying agents: “Personal Service Agents” and “Life Agents”.
- We found that people generally preferred to interact with personalizing agent identities that re-embody across services over service-specific personalizing agent identities and embodiment-specific identities.
- Our participants’ comments revealed affordances of co-embodying and personalizing agents that would provide value: emotional support and personality customization.
- We also identified concerns surrounding such agent behaviors, and possible ways to assuage those concerns:
  - People may worry about uncontrolled context-crossing of agent identities; therefore, this should be a toggle setting that users can control.
  - Lack of understanding of social context can bring about perceived and real personal privacy risks and awkwardness; therefore, co-embodying agent identities should follow social norms, legibly signal when they are accessing different people’s data and directing interaction to different users, and communicate their understanding of social context (i.e., the relationships among multiple simultaneous users).

- We found that people are uncomfortable with service agents that communicate with each other in humanlike ways when not directly responding to a user's request.
- We built a custom service robot that can be used in and modified for future service design research, as well as inspire the design of commercial service robots.
- We pioneered a new variation on the method of user enactments that emphasizes adding overall structure, comparisons, and events in order to better understand intermediate-level design knowledge (this is elaborated in Chapter 9).

Chapter 4: Comparing Personalized Interactions with Fluidly-Embodied Service Robots: Storyboards Study

- Our findings inspire specific design implications for creating appropriate robot identification and re-embodiment behaviors based on the service setting (see Table 4.4).
- This work contributes an example of how two exploratory studies that assess similar design concepts in vastly different ways can complement each other. In the previous study (Reig et al., 2020, described in Chapter 3), we exposed a smaller number of people to a richer set of immersive experiences, and collected detailed feedback. In this study, we collected data from a large number of participants based on a large number of low-fidelity stimuli deployed in a medium-scale online study. This allowed us to test specific questions and comparisons that arose in the first study.

Chapter 5: Perceptions of Multi-Robot Failure Recovery Strategies

- We found that trust and perceived competence of a multi-robot system were highest when a single robot with a single identity recovered on its own.
- We found that a single agent identity re-embodying into a new robot brought about higher perceptions of trust and competence following a failure than a second robot with a separate identity.
- We found that observers attribute failures that are recovered using re-embodiment to hardware problem more than they attribute failures that are recovered using a second robot (with a second agent identity) to a hardware problem.
- Our study suggests that after seeing a robot system experience a failure, people will be more likely to want to work with it again if they perceive it to have more agency.

Chapter 6: Agent Affiliation, Reference Cues, and Roles in Smart Environments

- We found that a singular agent was more likely to be perceived as "being" the environment than were domain- and user-affiliated agents.
- We found that an agent that spoke in third person was more likely to be perceived as "representing" the environment.

- We found evidence that a third person language perspective can lead to increased trust and perceptions of confidence over a first person perspective.
- Our findings suggest that in this kind of setting, agent affiliation does not impact trust or social perceptions of the agent or the environment.
- We designed the "Space Habitat Task", which can be used in future studies to simulate collaboration among humans and agents in smart environments.

Chapter 7: Perceptions of Agent Loyalty with Ancillary Users

- We found that when interacting with an agent identity that is not their own, people are likely to perceive that it is loyal to someone else (i.e., assume that it is acting in the interests of its primary user). This is especially true if it provides low-quality information about the primary user to the immediate user.
- We found that, across three distinct embodiments, low information quality negatively impacted the experience of collaborating with an agent and willingness to follow the agent's instructions.
- We designed the "Spy Task", which can be used in future studies to explore social variables and team constructs during complex, unstructured human-agent collaboration.

Chapter 8: Visions of Future Smart Homes

- The story set contained themes and possible end-user needs that can serve as food for thought for researchers and designers of smart home devices. These included:
  - how seamless interactions in future smart homes may reflect not just fulfillment but also anticipation of the needs of their inhabitants
  - people's tendency to implicitly and explicitly project class and gender onto smart home devices, sometimes in ways that mimic human social biases
  - that people do not readily imagine multi-user interactions, nor social interaction with smart home devices
  - that people may assume—but not necessarily accept—that seamless and integrated smart home interactions necessarily mean a lack of privacy
- This work provides an example of using the story completion method to explore possible futures of human-smart home interaction.

In addition, this thesis contributes two conceptual frameworks for the study of human–system interaction in socially and physically complex settings. One is motivated by a critical need (elucidated partly by the work in this thesis) to define and frame just what it means to study user experience in "smart environments". Grounded in historical definitions of "agents" and "smart environments", research into interface design and user values in smart homes, and theory surrounding human-robot teaming, the *five lenses to a systemic view of smart environments* model (Figure 9.1) articulates five distinct perspectives for researchers and designers to leverage.

This thesis is also broadly concerned with roles and opportunities for AI *agent identities* that are untethered to any particular embodiment, service, time, or place. Each

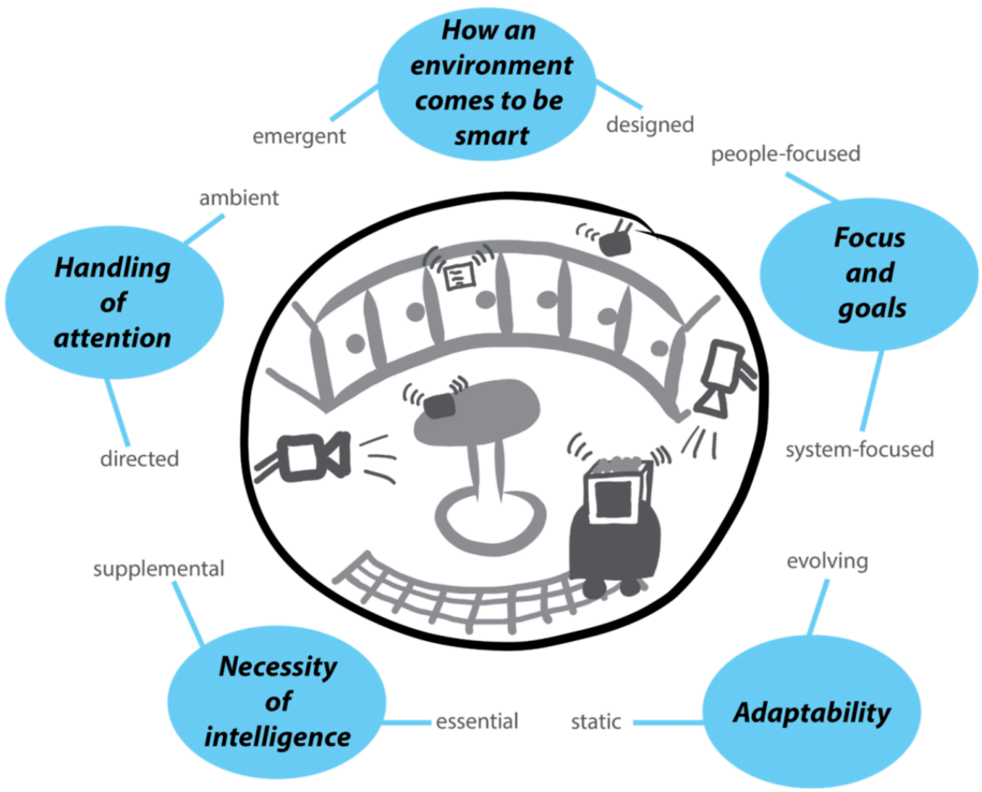

FIGURE 9.1: Five lenses for a systemic view of smart environments, with example variations on the lenses. This figure first appears in Chapter 6 and is repeated here for convenience.

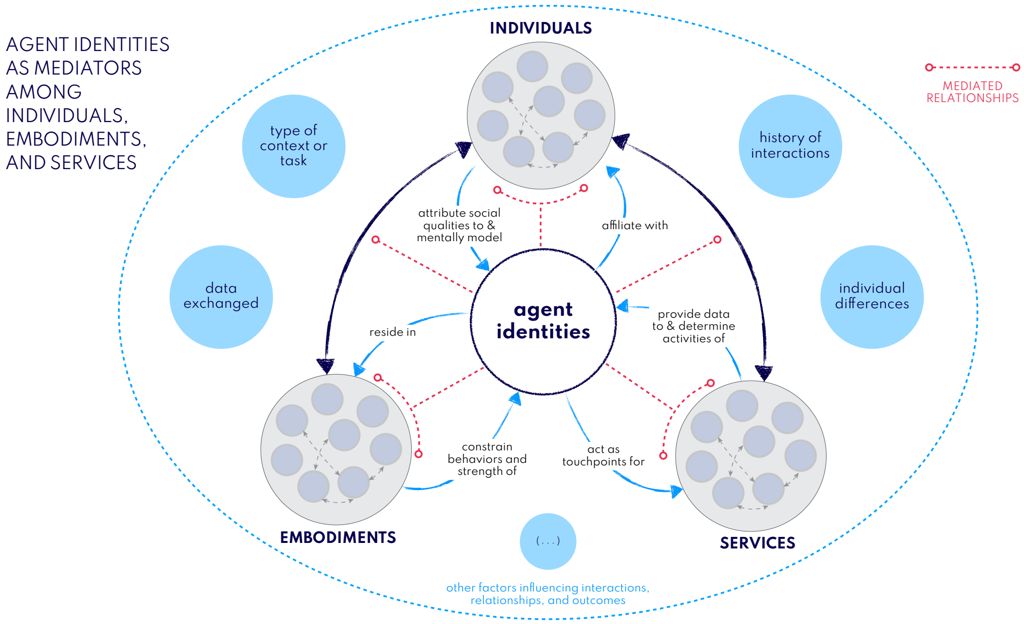

FIGURE 9.2: Conceptual model of agent identity as a mediator in relationships among individuals, embodiments, and services. This figure first appears in Chapter 1 and is repeated here for convenience.

chapter poses a set of research questions about agent identities (conceptualization, presentation to users, etc.) in multi-person, multi-robot, and/or multi-service relationships. Each chapter then explores one or more scenarios that highlight new ways in which a person, an agent identity, and some *other entity*—where the *other entity* is either another person, a service or company, a robot, or a (simulated) physical environment—might interact. The insights revealed by these explorations suggest that agent identities can be conceptualized as mediating entities among individuals, embodiments, and services[1]. The *agent identities as mediators* model (Figure 9.2) shows how the design and behavior of an agent identity that represents or associates with a person, embodiment, or service may—beyond impacting the way the agent itself is perceived—mediate relationships between end-users and these various affiliate objects and entities.

It is my hope that future research efforts will expand upon these frameworks as they are able to and reconsider them as they need to. Specific future research directions include:

- Identifying new lenses to the user experience of smart environments.
- Considering the "other factors" (e.g., actual interaction history, individual differences) in the *agent identities as mediators* model in rich detail, e.g., by theorizing about and validating conceptual models of interactions among individuals, embodiments, and services that center on those factors (instead of agent identities).
- Validating the relationships among the entities in the agent identities as mediators model, including investigating whether those relationships still take shape with

[1] According to STS definitions of mediating technologies; see Chapter 2.

non-conversational and non-human-like agents and robots, and whether they can be sustained over the long term.

9.2 Reflections on methodology

The research questions posed in this thesis targeted intermediate levels of design knowledge (see Edmondson and McManus, 2007) and previously unexplored forms of human-agent interaction. In order to address them, we developed a number of new stimuli, tasks, and methods. Many of these are discussed in detail in the chapters in which they are introduced. This section reiterates and expands on these methodological advancements.

9.2.1 Approach: structured user enactments

This section was previously published in a lightly peer-reviewed paper:

Samantha Reig, et al. (May 2020.) "Structured User Enactments to Evoke Future-Based Reflections on the Present". *CHI 2020* Workshop on Design Fictions.

When a design is brand new and implementations of it are still largely imaginary, researchers can rely on several forms of design fictions to study it from a future-oriented perspective. Different methods of "fiction" involve different activities: technology probes evoke initial responses to a new idea and facilitate communication between researchers and stakeholders (Boehner et al., 2007), props and dramas serve as "things to think with" in engaging people in participatory design (Brandt and Grunnet, 2000), etc. Their common thread is a focus on the future that is not bounded by the realities and possibilities of the present. A related method, User Enactments (Odom et al., 2012; Zimmerman and Forlizzi, 2017; Davidoff et al., 2007) (UE), exposes participants to low-fidelity mock-ups of future experiences with emerging technologies. Through an active walk-through of scenarios that incorporate interactions with and surrounding the novel technologies, participants get a brief feel for each of several "possible futures" (Zimmerman and Forlizzi, 2017), form impressions about their experiences, and reflect on how the designs might fit (or not fit) into their lives. This approach is best put to use when there exist general topics of interest that researchers want to explore, but no concrete foundation on which to base predictions. Enactments, like many of their methodological relatives, boast an ability to induce surprising responses and reveal other insights that designers would not have known to inquire about—the "unknown unknowns" of a new technology. Once enough is understood to craft high-fidelity prototypes, researchers can turn to more evaluative methods. When a product is close to realization or newly realized, researchers and developers may use discourse analysis to understand its place in the present (Harmon and Mazmanian, 2013), or concept videos to look to the near future (Wong and Mulligan, 2016). As technologies become real, research into how to improve their design continues in research and in practice through workshops, Wizard-of-Oz studies, and user testing.

For a portion of the lifespan of a new technology, designers know that the idea for the future technology is worthy of study, but do not yet know enough about how it will fit into the world to drill down on specific features of its visual, physical, or interaction design. During this window in time, the technology is not young enough for research that is purely-exploratory in nature, but it is also not mature enough for researchers to have sufficient knowledge to formulate testable hypotheses about it. Instead, research focused at this point in the development pipeline seeks to build, through exploration, a stronger foundation for understanding possibilities of the technology, thus adding more color to the picture of possible futures that involve it.

While the "standard" form of user enactments makes good use of low-fidelity prototypes to study early-stage ideas, it is limiting to mid-stage ideas: it emphasizes breadth, and in turn loses some of the richness that would come from modeling interactions surrounding the prototypes in enough depth to compare them. These interactions and the devices on which they depend do not yet exist, so a Wizard-of-Oz design is not believable. How, then, should we use research to better approach the question of how to "make the right thing" (Buxton, 2010; Zimmerman, Forlizzi, and Evenson, 2007) when an idea is in this middle stage of development? One option is to introduce some control over a user enactments study, thereby generating opportunities to draw comparisons between different versions of similar interactions.

In Chapter 3, we modified the user enactments method to include extra layers of structure in a qualitative study that we ran to probe the broad research question: *How should robots that personalize their service for multiple different users behave in public settings?* Since public robots are currently mostly limited to generic interactions, while personal devices in more private spaces are more likely to draw from personal data and hold person-specific "conversations", we knew this question lent itself to a qualitative and open-ended exploration. A method that allowed people to sample "a menu of possible futures" (Zimmerman and Forlizzi, 2017) would help participants envision many areas of the very broad space of human-robot service interactions-to-be while also giving them concrete personal experiences to anchor their reflections.

We began by brainstorming (1) several physical settings in which this sort of interaction might occur and (2) several different framings for interactions with personalizing agents embodied in public robots (e.g., a user's own personal agent vs. a service-maintained agent that has personal history with a user). As is common in the early stages of many exploratory design research approaches, our goal was to first collect as many ideas as we could, and then narrow down our list to the set of scenarios in which we could most easily and compellingly situate low-fidelity prototypes of rich future interactions. Throughout the scenario development process, we returned to our early notes numerous times to identify specific areas of interest touched upon by our guiding research question. We also acted out—first via improvisation, and then guided by our evolving scripts—several possible scenarios ourselves, and in doing so, varied individual events within the enactments to see how such variations might impact eventual participants' opportunities to think about and respond to a diverse set of experiences. We began collecting and creating props and environments to use to facilitate rich, but obviously artificial, service interactions.

As our study design took shape, we realized that several issues we wanted to probe lent themselves to a more "controlled" exploration than is traditionally accessible by open-ended user enactments. At this point, we had already selected domains in which to construct scenarios: shopping, an everyday activity that could perhaps be made more interesting or more efficient by personalized recommendations; health, a more serious

Domain area	The broad context for the design investigation. This can be narrow and place-focused, such as *smart bedroom*, as in Odom et al., 2012 and Odom et al., 2014, or focused more generally on a type of interaction, such as *transactional service interactions*, as in Luria et al., 2019 and Reig et al., 2020.
Future technology	The design that does not yet exist, but soon might. The idea(s) can be a general and realizable in several forms, like *robot re-embodiment* (Luria et al., 2019) or *smart wall*, or specific, like *Status Quilt* (Odom et al., 2012). Participants interact with prototypes of these design concepts.
Scenarios	Prompts or scripts that inspire and/or scaffold participants' experiences with different versions of the future technologies. They establish scenarios that are generally flexible, but contain deliberately-placed *events* ("plot points" that always happen) that address key themes.
Features of interest	The aspects of scenarios that, when varied and/or combined in deliberate ways, comprise the "structure" of the enactments study.

TABLE 9.1: Elements of structured user enactments.

topic that can involve emotional sensitivity and make privacy of particular concern; and travel, a setting that both evokes novelty and can benefit from some degree of familiarity. We had also designed three framings for personalized interactions with service robots: one service agent that interacted with everyone (the present-day or "baseline" design), multiple service agents that each interacted with particular people, and multiple service agents that were actually owned by the user and migrated (see Luria et al., 2019) into different robots to provide different services.

We wanted to keep some things consistent within each setting and framing to see how participants' impressions generalized: for example, with company-owned, person-specific agents, we always included one exchange in which the two voices said the exact same thing at the exact same time. On the other side of the coin, certain interesting events were only testable in specific combinations of settings and interaction framings: for example, in a health clinic lobby, a user's own agent could make reference to how much water the user had had to drink that morning—but this sort of interaction would make much less sense if it were to take place in a department store instead of in a medical setting, or if the agent were not already the user's companion.

We therefore developed nine total scripts, each one involving one setting and one framing. Some interactions were written into all three scripts of the same framing or setting, while others only occurred in one of the nine scripts. An experimenter controlled the prototypes according to the scripts, but deviated from the script as needed. Certain segments were written to make it hard for participants to respond in novel ways; we wanted people to say and hear specific things that would bring about specific experiences that we thought were representative of a phenomena we wanted to probe. This deliberate, selective imposition of non-negotiable events is another way in which structured UEs differ from standard ones. The resulting design can be conceptualized similarly to a 3 x 3 study—each of two key concepts manifests three different ways, like three levels of two variables—but there are no distinct outcome measures, and the variables of interest are not meant to be studied in isolation, the experiment is not controlled, and the study is still treated as an exploratory method.

It can be challenging to close in on fruitful ways to investigate design ideas that are beyond initial inception but still immature. Adding structure to scenario-based exploratory research is one way to address this challenge. Running a controlled study (e.g., a Wizard-of-Oz scenario in which a robot always behaves the same way except in terms of key variables) can artificially constrain the research to a smaller subset of the design space than researchers want to explore. Relying wholly on emergent design fictions (e.g., allowing the scenarios to take any number of different narrative branches each time they are run) does not give researchers sufficient opportunity to observe differences in responses based on deliberate inconsistencies that they strongly suspect will make for interesting comparisons.

Chapter 3 revealed tradeoffs between different ways of employing elements of fictionalization in the research of future interactions. For a Wizard-of-Oz study to be effective, the technology needs to be plausible enough that the possibility of participants seeing through the veil poses minimal risk to the research outcomes. With futuristic enactments, control is sacrificed, but immediate plausibility is not required; instead, there is no illusion, and participants knowingly suspend their disbelief. We leaned on the flexible nature of exploratory research and combined these different approaches into the hybrid, open-ended-yet-constrained, method of structured user enactments. This method allows participants to be contributors to the fiction of the future and the "subjects" of the study (as their responses are observed by the researchers) at the same time. In both of these roles, they can inform design by helping researchers decide whether and how to move forward with early-stage ideas about new kinds of interactions. It is our hope that through future research and continued discussion with the HCI research community, we can position this new approach to a known method among other accounts of novel uses of methodologies that draw from design fiction.

9.2.2 Study tasks

Behavioral research in HRI often makes use of established study tasks from organizational, social, and moral psychology, and from related bodies of literature. Two novel study tasks were introduced in this thesis:

Ch. 7 The *Spy Task* is an escape room-like game in which a participant collaborates with an agent or robot to complete a word puzzle. It is designed to be played in-person. The participant searches around the room for physical objects that contain symbols that can be identified or translated using a reference key (we used flat Braille letters). Certain objects are "decoys" which, if found, penalize (or appear to penalize) the score. The agent provides clues about where the objects are located; some such "clues" actually lead to decoys. When the participant has collected and identified all of the objects, they solve an anagram, which is the answer to the puzzle and signals successful completion of the task.

Ch. 6 The *Space Habitat Task* is a game consisting of sequential puzzles and designed to be played in a browser. As a participant clicks through several screens, they look up information, perform logical operations, and input values to solve the puzzles.

A chatbot assists the participant with the puzzles, providing guidance about what to do with interface elements and needed numerical and textual values. Each puzzle requires information that is only known by the agent(s), information that is only known by the participant(s), and interface interaction by the participant (e.g., to type or click the answer). Successful completion of each puzzle (i.e., subtask) is required to advance to the next puzzle.

These tasks can be employed by future researchers seeking ways to (1) simulate interaction between people and socially interactive agents in unstructured, time-sensitive settings, particularly when the agent's performance is flawed (Spy Task) and/or (2) explore collaboration among humans and chatbots in smart environments (Space Habitat Task).

9.2.3 Research systems and tools

This work also generated a number of tangible outputs in the form of systems, stimuli, artifacts, and research tools. Here is a list of such outputs:

Ch. 3 Interaction scripts for nine scenarios (hotel, department store, health clinic × singular agent, personal service agents, life agents) involving robot re-embodiment, co-embodiment, and personalization in service interactions. Future research could adapt these scripts for use in studies on re-/co-embodiment and personalization, studies that employ the Structured User Enactments method, and studies that focus on human-robot interaction in the three environments we prototyped.

Ch. 3 A custom service robot built on top of an iRobot Create base.

Ch. 3 Protocol and "wizard interface" for operating the robot in the context of the study (see `https://github.com/CMU-TBD/HRI20-Not-Some-Random-Agent-Personalizing-Service-Robot`).

Ch. 4 A set of storyboards (illustrations and captions) depicting HAI with re-embodying expert agents and robots across different service domains and different roles within the same service domains.

Ch. 6 A fully developed website that provides the structure and interface for a sequential puzzle-style human-agent collaboration game. In this thesis, it was used for the Space Habitat Task. In future work, it could be modified to study similar research questions (related to agent design in smart environments) in different contexts: for example, the basic structure of the game could be maintained, but the narrative and puzzles revised to reflect collaboration in a smart home, healthcare, or transportation setting.

Ch. 6 A Dialogflow chatbot trained on the puzzles and context in the Space Habitat Task.

Ch. 6 A guide for implementing the NASA-TLX in a Qualtrics survey (see `https://github.com/CMU-TBD/qualtrics-tlx`).

9.3 Design implications

The findings in this thesis lend themselves to a number of design implications. Our work on service design brings about recommendations for the design of service robots. Our work on multi-robot failures and ambiguous agent intent has implications for the design collaborative systems. Several chapters surface considerations for design of smart environments broadly, as well as for commercial devices in homes and other smart environments. For easy reference, the recommendations made throughout the thesis are compiled here.

9.3.1 For service robots

Ch. 3 Co-embodiment as a way of performing service should be "opt-in": Robots in public settings can enable co-embodiment, but should not be embodied by two agents at the same time by default.

Ch. 3 Robots that use users' personal data should explicitly signal when they do and do not have access to individuals' data.

Ch. 3 When perceived risk is high (as in a medical setting), robot design should prioritize communication of expertise over personalization.

Ch. 3 Whenever possible, designers of re-embodyable robots should provide a means of continually determining the presence and absence of different agents.

Ch. 4 If personalization is required for the service, a robot should identify the user. Otherwise, service robots should avoid identifying people.

Ch. 4 Service robots that identify customers should not do so using facial recognition.

Ch. 4 In low-risk service contexts, robots can re-embody to provide a better service experience.

Ch. 4 If robots fulfill several tasks in one larger domain, re-embodiment can improve the service experience. A robot should not re-embody to do tasks in different domains.

Ch. 7 When designing an agent for private and public settings, to increase trust in and perceived loyalty of the agent, consider incorporating behavior that leads people to view it as a good teammate.

9.3.2 For human-agent collaboration

Ch. 7 When an agent interacts under conditions of uncertainty, it should forewarn the human partner that it may make an error.

Ch. 5 Robots that work closely with humans in task-oriented settings might benefit from being designed to take on a social "software identity" that can persist across embodiments to maintain trust after unexpected errors and failures.

Ch. 6 To appropriately calibrate trust and expectations, designers of conversational agents and robots that use natural language should consider having them use third person perspective when confidence is high and first person perspective when confidence is low.

Ch. 6 To reduce mental demand, use a singular agent when tasks or "jobs" are closely related, and multiple agents when they are not.

9.3.3 For smart home devices and smart environments

Ch. 6 The study of human-AI interaction in complex settings can benefit from a focus on the smart environment as a unit of analysis.

Ch. 6 Researchers and designers should consider several perspectives in creating and understanding UX in smart environments: how an environment comes to be smart, focus and goals, adaptability, necessity of intelligence, and handling of attention.

Ch. 6 If a smart environment is intended to be seen as a unified entity, then user experience within it should involve a single interactive agent or intelligence.

Ch. 8 In order to be adopted, smart home devices need to be transparent about what data is being collected when, and with whom it is being shared.

Ch. 8 To support diverse users' goals for AI that supports everyday tasks in the home, smart home devices should be designed to support improvisation and skill-building together with technology.

Ch. 8 Future smart homes should include AI and learning in order for systems to become familiar enough with users' habits to anticipate their needs and wants.

Ch. 8 Future smart homes should be designed with a service layer and an experience layer in mind for overarching coordination and seamless interaction.

While these recommendations are phrased conclusively, I emphasize that they are not absolutes. They are derived primarily from exploratory work, and the claims they make have not been thoroughly validated. The findings that inspired such recommendations may have, in many cases, been influenced by current events, evolving social norms, and portrayals of agent identities in popular media and science fiction; all of these factors are products of a particular time, and can change over time.

9.4 Closing remarks

This dissertation thoroughly explored and deeply analyzed the design concept of robot re-embodiment in terms of its potential value, perceived risks, and ethical considerations. More broadly, we illustrated several ways in which interaction design focused specifically on the concept of AI agent *identities*—sometimes, but not always, embedded in physical robots—can be meaningful in shaping relationships among people, services, and physical embodiments (i.e., hardware). In the process, we combined behavioral and design and quantitative and qualitative approaches; pioneered a new variation on an established research-through-design method; created several prototypes of future robots, agents, and environments; and developed new complex human-agent collaboration tasks. We also spoke with dozens of people face-to-face about their beliefs, dreams, goals, and fears surrounding future agent technologies—and we read, through online studies, the remarks of hundreds more.

How human-agent interaction should evolve as robots and other socially interactive systems become increasingly prevalent, competent, and integrated is one of the "big questions" of our time. The research problems that I and my collaborators have explored—and the mixed approaches we have taken to exploring them—provide examples of how this complex and at times philosophical topic can be broken down into its parts and empirically examined. Our contributions demonstrate and emphasize the importance of considering a given system's human-like *and* non-human-like social behavior (e.g., how it leverages personal data when interacting with users, with whom or what it presents itself as being affiliated) distinctly from its physical design and with explicit attention to its context. It is my hope that this thesis provides a strong foundation for continual scientific discourse on when, how, and why we should (or should not) design software agent identities that persist across embodiments, services, and time.

Bibliography

Abdi, Hervé (2003). "Factor rotations in factor analyses". In: *Encyclopedia for Research Methods for the Social Sciences. Sage: Thousand Oaks, CA*, pp. 792–795.

Abramovich, Giselle (2018). *Study Finds Consumers Are Embracing Voice Services. Heres How*. en. URL: https://blog.adobe.com/en/publish/2018/09/06/adobe-2018-consumer-voice-survey (visited on 09/05/2020).

Admoni, Henny, Christopher Datsikas, and Brian Scassellati (2014). "Speech and gaze conflicts in collaborative human-robot interactions". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 36. 36.

Admoni, Henny and Brian Scassellati (May 2017). "Social Eye Gaze in Human-robot Interaction: A Review". In: *J. Hum.-Robot Interact.* 6.1, pp. 25–63. ISSN: 2163-0364. DOI: 10.5898/JHRI.6.1.Admoni. URL: https://doi-org.proxy.library.cmu.edu/10.5898/JHRI.6.1.Admoni.

Admoni, Henny et al. (2016). "Robot Nonverbal Behavior Improves Task Performance In Difficult Collaborations". In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. HRI '16. Christchurch, New Zealand: IEEE Press, pp. 51–58. ISBN: 978-1-4673-8370-7.

Aibo (n.d.). URL: https://us.aibo.com/.

Auxier, Brooke (2020). *5 things to know about Americans and their smart speakers*. en-US. URL: https://www.pewresearch.org/fact-tank/2019/11/21/5-things-to-know-about-americans-and-their-smart-speakers/ (visited on 09/05/2020).

Baldwin, Roberto (2019). *Alexa can mimic celebrities, starting with Samuel L. Jackson*. Engadget. URL: https://www.engadget.com/2019-09-25-amazon-alexa-celebrity-voices.html.

Ball, Dwayne, Pedro S Coelho, and Manuel J Vilares (2006). "Service personalization and loyalty". In: *Journal of Services Marketing* 20.6, pp. 391–403.

Banisetty, Santosh Balajee and Tom Williams (2021). "Implicit communication through social distancing: Can social navigation communicate social norms?" In: *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 499–504.

Bartneck, Christoph et al. (2009a). "Does the design of a robot influence its animacy and perceived intelligence?" In: *International Journal of Social Robotics* 1, pp. 195–204.

Bartneck, Christoph et al. (Jan. 2009b). "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots". en. In: *International Journal of Social Robotics* 1.1, pp. 71–81. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-008-0001-3. URL: http://link.springer.com/10.1007/s12369-008-0001-3 (visited on 04/24/2018).

Bejarano, Alexandra et al. (2022). “You had me at hello: The impact of robot group presentation strategies on mental model formation”. In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 363–371.

Belk, Russell W (1988). “Possessions and the extended self”. In: *Journal of consumer research* 15.2, pp. 139–168.

— (2013). “Extended self in a digital world”. In: *Journal of consumer research* 40.3, pp. 477–500.

Bell, Chip R and Ron E Zemke (1987). “Service breakdown: the road to recovery”. In: *Management review* 76.10, p. 32.

Bellucci, Andrea et al. (2019). “Augmenting objects at home through programmable sensor tokens: A design journey”. In: *International Journal of Human-Computer Studies* 122, pp. 211–231.

Bergh, Don van den et al. (2020). “A tutorial on conducting and interpreting a Bayesian ANOVA in JASP”. In: *LAnnee psychologique* 120.1, pp. 73–96.

Beyer, Hugh and Karen Holtzblatt (1997). *Contextual design: A customer-centered approach to systems designs*.

Blom, Jan O and Andrew F Monk (2003). “Theory of personalization of appearance: why users personalize their PCs and mobile phones”. In: *Human-Computer Interaction* 18.3, pp. 193–228.

Bodala, Indu P et al. (2020). “Modeling the interplay of trust and attention in hri: An autonomous vehicle study”. In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 145–147.

Boehner, Kirsten et al. (2007). “How HCI interprets the probes”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1077–1086.

Bohn, Dieter (2019). *Amazon creates a huge alliance to demand voice assistant compatibility*. The Verge.

Brandt, Eva and Camilla Grunnet (2000). “Evoking the future: Drama and props in user centered design”. In: *Proceedings of Participatory Design Conference (PDC 2000)*, pp. 11–20.

Branham, Stacy M., Shahtab Wahid, and D. Scott McCrickard (2007). “Channeling creativity: Using storyboards and claims to encourage collaborative design”. In: *Workshop on Tools in Support of Creative Collaboration (part of Creativity & Cognition 2007)*, pp. 1–4.

Braun, Virginia and Victoria Clarke (Jan. 2006). “Using Thematic Analysis in Psychology”. In: *Qualitative Research in Psychology* 3.2, pp. 77–101. ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa.

— (2013). *Successful Qualitative Research: A Practical Guide for Beginners*. Thousand Oaks, CA: sage.

Braun, Virginia et al. (2019). “Qualitative Story Completion: Possibilities and Potential Pitfalls”. In: *Qualitative Research in Psychology* 16.1, pp. 136–155.

Brooks, Rodney A (1991). “Intelligence without representation”. In: *Artificial intelligence* 47.1-3, pp. 139–159.

Bršćić, Dražen et al. (2015). “Escaping from children’s abuse of social robots”. In: *Proceedings of the tenth annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 59–66.

Buchenau, Marion and Jane Fulton Suri (Aug. 2000). "Experience prototyping". In: *Proceedings of the 3rd conference on Designing Interactive Systems: processes, practices, methods, and techniques*. DIS '00. New York, NY, USA: Association for Computing Machinery, pp. 424–433. ISBN: 978-1-58113-219-9. DOI: 10.1145/347642.347802. URL: https://doi.org/10.1145/347642.347802 (visited on 02/04/2021).

Buxton, Bill (2010). *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann.

Cambre, Julia et al. (2020). "'All Rise for the AI Director': Eliciting Possible Futures of Voice Technology through Story Completion". In: *Proceedings of the 2020 ACM designing interactive systems conference*, pp. 2051–2064.

Carpinella, Colleen M. et al. (2017). "The Robotic Social Attributes Scale (RoSAS): Development and Validation". In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. Vienna, Austria: ACM, pp. 254–262. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020208. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/2909824.3020208.

Cassell, J. et al. (1999). "Embodiment in Conversational Interfaces: Rea". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: ACM, pp. 520–527. ISBN: 0-201-48559-1. DOI: 10.1145/302979.303150. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/302979.303150.

Cassell, Justine (Oct. 2009). "Getting by with a Little Help from Your (Virtual) Friends". In: *J. Comput. Sci. Coll.* 25.1, pp. 7–7. ISSN: 1937-4771. URL: http://dl.acm.org.proxy.library.cmu.edu/citation.cfm?id=1619221.1619223.

Cassell, Justine et al. (2000). *Embodied Conversational Agents*. Cambridge, MA, USA: MIT Press. ISBN: 0-262-03278-3.

Chaves, Ana Paula and Marco Aurelio Gerosa (2018). "Single or Multiple Conversational Agents? An Interactional Coherence Comparison". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

Chung, Jen Jen et al. (2019). "The Impact of Agent Definitions and Interactions on Multiagent Learning for Coordination". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. Montreal QC, Canada: International Foundation for Autonomous Agents and Multiagent Systems, pp. 17521760. ISBN: 9781450363099.

Coeckelbergh, Mark (2012). "Can we trust robots?" In: *Ethics and information technology* 14, pp. 53–60.

Connectivity Standards Alliance (2022). *Matter: The Foundation for Connected Things*.

Coradeschi, Silvia et al. (2011). "Social Robotic Telepresence". In: *Proceedings of the 6th International Conference on Human-robot Interaction*. HRI '11. Lausanne, Switzerland: ACM, pp. 5–6. ISBN: 978-1-4503-0561-7. DOI: 10.1145/1957656.1957660. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/1957656.1957660.

Correia, Filipa et al. (2019). "Exploring prosociality in human-robot teams". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 143–151.

Cushing, Amber L (2011). "Self extension and the desire to preserve digital possessions". In: *Proceedings of the American Society for Information Science and Technology* 48.1, pp. 1–3.

Das, S.K. and D.J. Cook (June 2006). "Designing and modeling smart environments". In: *2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks(WoWMoM'06)*, 5 pp.–494. DOI: 10.1109/WOWMOM.2006.35.

Davidoff, Scott et al. (2007). "Rapidly exploring application design through speed dating". In: *International Conference on Ubiquitous Computing*. Springer, pp. 429–446.

Deng, Eric, Bilge Mutlu, and Maja J Mataric (2019). "Embodiment in socially interactive robots". In: *Foundations and Trends® in Robotics* 7.4, pp. 251–356.

Desai, Munjal (2007). "Sliding scale autonomy and trust in human-robot interaction". PhD thesis. University of Massachusetts. Lowell.

Desai, Munjal et al. (2009). "Creating trustworthy robots: Lessons and inspirations from automated systems". In: *Proceedings of the AISB Convention: New frontiers in human-robot interaction*, pp. 6–9.

Desai, Munjal et al. (Mar. 2012). "Effects of changing reliability on trust of robot systems". In: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. HRI '12. New York, NY, USA: Association for Computing Machinery, pp. 73–80. ISBN: 978-1-4503-1063-5. DOI: 10.1145/2157689.2157702. URL: https://doi.org/10.1145/2157689.2157702 (visited on 09/27/2020).

Desai, Munjal et al. (Mar. 2013). "Impact of robot failures and feedback on real-time trust". en. In: IEEE, pp. 251–258. ISBN: 978-1-4673-3101-2 978-1-4673-3099-2 978-1-4673-3100-5. DOI: 10.1109/HRI.2013.6483596. URL: http://ieeexplore.ieee.org/document/6483596/ (visited on 05/24/2018).

Dienes, Zoltan and Neil Mclatchie (2018). "Four reasons to prefer Bayesian analyses over significance testing". In: *Psychonomic bulletin & review* 25, pp. 207–218.

Dix, Alan (2007). "Designing for appropriation". In: *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21*, pp. 1–4.

Duffy, Brian R. et al. (2003). "Agent chameleons: Agent minds and bodies". In: *Proceedings 11th IEEE International Workshop on Program Comprehension*. IEEE, pp. 118–125.

Edison Research (2022). *The Smart Audio Report 2022 from NPR and Edison Research*. URL: https://www.edisonresearch.com/smart-audio-report-2022-from-npr-and-edison-research/ (visited on 06/28/2022).

Edmondson, Amy C and Stacy E McManus (2007). "Methodological fit in management field research". In: *Academy of management review* 32.4, pp. 1246–1264.

Eyal, Peer et al. (2021). "Data quality of platforms and panels for online behavioral research". In: *Behavior Research Methods*, pp. 1–20.

Ezer, Neta (2008). "Is a robot an appliance, teammate, or friend? Age-related differences in expectations of and attitudes toward personal home-based robots". PhD Thesis. Georgia Institute of Technology.

Fallon, Corey K. et al. (May 2010). "The calibration of trust in an automated system: A sensemaking process". In: *2010 International Symposium on Collaborative Technologies and Systems*, pp. 390–395. DOI: 10.1109/CTS.2010.5478488.

Floridi, Luciano and Jeff W Sanders (2004). "On the morality of artificial agents". In: *Minds and machines* 14.3, pp. 349–379.

Fong, Terrence (2001). "Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation". en. In: *Technical Report CMU-RI-TR-01-34.*, p. 198.

Forlizzi, Jodi (2007). "How robotic products become social products: an ethnographic study of cleaning in the home". In: *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pp. 129–136.

— (2008). “The product ecology: Understanding social product use and supporting design culture”. In: *International Journal of design* 2.1.

Franklin, Stan and Art Graesser (1997). “Is it an Agent, or Just a Program?: A taxonomy for autonomous agents”. en. In: *Intelligent Agents III Agent Theories, Architectures, and Languages*. Ed. by J. G. Carbonell et al. Vol. 1193. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 21–35. ISBN: 978-3-540-62507-0 978-3-540-68057-4. DOI: 10.1007/BFb0013570. URL: http://link.springer.com/10.1007/BFb0013570 (visited on 06/22/2020).

Fraune, Marlena R., Selma Šabanović, and Eliot R. Smith (2017). “Teammates First: Favoring Ingroup Robots Over Outgroup Humans”. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 1432–1437.

Fraune, Marlena R. et al. (2017). “Threatening Flocks and Mindful Snowflakes: How Group Entitativity Affects Perceptions of Robots”. In: *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 205–213.

Fraune, Marlena R. et al. (2020). “Effects of Robot-Human Versus Robot-Robot Behavior and Entitativity on Anthropomorphism and Willingness to Interact”. In: *Computers in Human Behavior* 105. Publisher: Elsevier, p. 106220.

Furr, Mike (2011). *Scale construction and psychometrics for social and personality psychology*. SAGE Publications Ltd.

Fussell, Susan R. et al. (2008). “How People Anthropomorphize Robots”. In: *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. HRI ’08. Amsterdam, The Netherlands: ACM, pp. 145–152. ISBN: 978-1-60558-017-3. DOI: 10.1145/1349822.1349842. URL: http://doi.acm.org/10.1145/1349822.1349842.

Gabrecht, Katharina M (2016). “Human factors of semi-autonomous robots for urban search and rescue”. PhD thesis. University of Nottingham.

Gervits, Felix, Terry W. Fong, and Matthias Scheutz (2018). “Shared Mental Models to Support Distributed Human-Robot Teaming in Space”. In: *2018 AIAA SPACE and Astronautics Forum and Exposition*, p. 5340.

Gervits, Felix et al. (2020). “Toward Genuine Robot Teammates: Improving Human-Robot Team Performance Using Robot Shared Mental Models”. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 429–437.

Gillet, Sarah et al. (2021). “Robot Gaze Can Mediate Participation Imbalance in Groups with Different Skill Levels”. In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 303–311.

Gomez, Carles et al. (Jan. 2019). “Internet of Things for enabling smart environments: A technology-centric perspective”. en. In: *Journal of Ambient Intelligence and Smart Environments* 11.1, pp. 23–43. ISSN: 18761372, 18761364. DOI: 10.3233/AIS-180509. URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AIS-180509 (visited on 06/19/2020).

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann Jr. (2003). “A very brief measure of the Big-Five personality domains”. In: *Journal of Research in personality* 37.6, pp. 504–528.

Gosling, Samuel D et al. (2002). “A room with a cue: personality judgments based on offices and bedrooms.” In: *Journal of personality and social psychology* 82.3, p. 379.

Gravett, Karen (2019). “Story completion: Storying as a method of meaning-making and discursive discovery”. In: *International Journal of Qualitative Methods* 18, p. 1609406919893155.

Gravett, Karen and Naomi E. Winstone (2019). “Storying students becomings into and through higher education”. In: *Studies in Higher Education* 0.0, pp. 1–12. DOI: 10.1080/03075079.2019.1695112. URL: https://doi.org/10.1080/03075079.2019.1695112.

Greenleaf, Eric A. (1992). “Measuring extreme response style”. In: *Public Opinion Quarterly* 56.3. Publisher: Oxford University Press, pp. 328–351.

Groom, Victoria Jane (2010). *Self Extension Into Robots: An Examination Of Variables That Promote Overlap In The Concepts Of Self And Robot*. Stanford University.

Harmon, Ellie and Melissa Mazmanian (2013). “Stories of the Smartphone in everyday discourse: conflict, tension & instability”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1051–1060.

Hart, Sandra G and Lowell E Staveland (1988). “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research”. In: *Advances in psychology*. Vol. 52. Elsevier, pp. 139–183.

Hoffman, Guy and Xuan Zhao (2020). “A primer for conducting experiments in human–robot interaction”. In: *ACM Transactions on Human-Robot Interaction (THRI)* 10.1, pp. 1–31.

Hoffman, Guy et al. (2015). “Robot Presence and Human Honesty: Experimental Evidence”. In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’15. Portland, Oregon, USA: ACM, pp. 181–188. ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696487. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/2696454.2696487.

Horvitz, Eric (1999). “Principles of mixed-initiative user interfaces”. en. In: *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI ’99*. Pittsburgh, Pennsylvania, United States: ACM Press, pp. 159–166. ISBN: 978-0-201-48559-2. DOI: 10.1145/302979.303030. URL: http://portal.acm.org/citation.cfm?doid=302979.303030 (visited on 06/18/2020).

Huang, Chien-Ming and Bilge Mutlu (June 2013). “The Repertoire of Robot Behavior: Enabling Robots to Achieve Interaction Goals Through Social Behavior”. In: *J. Hum.-Robot Interact.* 2.2, pp. 80–102. ISSN: 2163-0364. DOI: 10.5898/JHRI.2.2.Huang. URL: https://doi-org.proxy.library.cmu.edu/10.5898/JHRI.2.2.Huang.

Höök, Kristina and Jonas Löwgren (2012). “Strong concepts: Intermediate-level knowledge in interaction design research”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 19.3. Publisher: ACM New York, NY, USA, pp. 1–18.

Ihde, Don (1990). “Technology and the lifeworld: From garden to earth”. In.

Isbister, Katherine and Clifford Nass (2000). “Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics”. In: *International Journal of Human-Computer Studies* 53.2, pp. 251–267.

Jackson, Ryan Blake and Tom Williams (2019). “On perceived social and moral agency in natural language capable robots”. In: *2019 HRI Workshop on The Dark Side of Human-Robot Interaction. Jackson, RB, and Williams*, pp. 401–410.

Jackson, Ryan Blake et al. (2021). “Design, Performance, and Perception of Robot Identity”. In: *Workshop on Robo-Identity: Artificial identity and multi-embodiment at HRI 2021*.

Jaschinski, Christina (Sept. 2014). “Ambient assisted living: towards a model of technology adoption and use among elderly users”. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. UbiComp ’14 Adjunct. New York, NY, USA: Association for Computing Machinery, pp. 319–324. ISBN: 978-1-4503-3047-3. DOI: 10.1145/2638728.2638838. URL: http://doi.org/10.1145/2638728.2638838 (visited on 08/13/2020).

Jennings, Nicholas R., Katia Sycara, and Michael Wooldridge (Mar. 1998). “A Roadmap of Agent Research and Development”. en. In: *Autonomous Agents and Multi-Agent Systems* 1.1, pp. 7–38. ISSN: 1573-7454. DOI: 10.1023/A:1010090405266. URL: https://doi.org/10.1023/A:1010090405266 (visited on 06/18/2020).

Jian, Jiun-Yin, Ann M. Bisantz, and Colin G. Drury (Mar. 2000). “Foundations for an Empirically Determined Scale of Trust in Automated Systems”. In: *International Journal of Cognitive Ergonomics* 4.1, pp. 53–71. ISSN: 1088-6362. DOI: 10.1207/S15327566IJCE0401_04. URL: https://doi.org/10.1207/S15327566IJCE0401_04 (visited on 09/22/2020).

Johnson, Matthew et al. (Mar. 2014). “Coactive Design: Designing Support for Interdependence in Joint Activity”. en. In: *Journal of Human-Robot Interaction* 3.1, p. 43. ISSN: 2163-0364. DOI: 10.5898/JHRI.3.1.Johnson. URL: http://dl.acm.org/citation.cfm?id=3109825 (visited on 06/11/2020).

Keijsers, Merel and Christoph Bartneck (2018). “Mindless robots get bullied”. In: *Proceedings of the 2018 acm/ieee international conference on human-robot interaction*, pp. 205–214.

Keysers, Christian, Valeria Gazzola, and Eric-Jan Wagenmakers (2020). “Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence”. In: *Nature neuroscience* 23.7, pp. 788–799.

Kiesler, Sara et al. (2008). “Anthropomorphic interactions with a robot and robot–like agent”. In: *Social Cognition* 26.2, pp. 169–181.

Kiesler, Tina and Sara Kiesler (2005). “My pet rock and me: An experimental exploration of the self extension concept”. In: *ACR North American Advances*.

Kim, Chyon Hae et al. (2013). “Recognition for psychological boundary of robot”. In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 161–162.

Kim, Taemie and Pamela Hinds (Sept. 2006). “Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction”. en. In: *RO-MAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*. Univ. of Hertfordshire, Hatfield, UK: IEEE, pp. 80–85. ISBN: 978-1-4244-0564-0 978-1-4244-0565-7. DOI: 10.1109/ROMAN.2006.314398. URL: http://ieeexplore.ieee.org/document/4107789/ (visited on 09/22/2020).

Kinsella, Bret (Apr. 2020). *Nearly 90 Million U.S. Adults Have Smart Speakers, Adoption Now Exceeds One-Third of Consumers*. en-US. Section: smart speaker. URL: https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/ (visited on 09/05/2020).

Knepper, Ross A. et al. (2015). “Recovering from failure by asking for help”. In: *Autonomous Robots* 39.3. Publisher: Springer, pp. 347–362.

Koay, Kheng Lee et al. (2009). "A user study on visualization of agent migration between two companion robots". In: *In Thirteenth International Conference on Human Computer Interaction*.

Koay, Kheng Lee et al. (2016). "Prototyping realistic long-term human-robot interaction for the study of agent migration". In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 809–816.

Kok, Bing Cai and Harold Soh (2020). "Trust in robots: Challenges and opportunities". In: *Current Robotics Reports* 1, pp. 297–309.

Kozak, Megan N., Abigail A. Marsh, and Daniel M. Wegner (2006). "What do i think you're doing? Action identification and mind attribution." en. In: *Journal of Personality and Social Psychology* 90.4, pp. 543–555. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.90.4.543. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.90.4.543 (visited on 06/16/2020).

Kriegel, Michael et al. (2011). "Robots meet IVAs: a mind-body interface for migrating artificial intelligent agents". In: *International Workshop on Intelligent Virtual Agents*. Springer, pp. 282–295.

Krueger, Joachim I and Patrick R Heck (2017). "The heuristic value of p in inductive statistical inference". In: *Frontiers in Psychology* 8, p. 908.

Kruschke, John K (2010). "What to believe: Bayesian methods for data analysis". In: *Trends in cognitive sciences* 14.7, pp. 293–300.

Kubo, Akinori (2010). "Technology as Mediation: On the Process of Engineering and Living with the" AIBO" Robot". In: *Japanese Review of Cultural Anthropology* 11, pp. 103–123.

Kwon, Minae, Sandy H. Huang, and Anca D. Dragan (2018a). "Expressing robot incapability". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 87–95.

Kwon, Minae, Sandy H Huang, and Anca D Dragan (2018b). "Expressing robot incapability". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 87–95.

Latour, Bruno (1999). *Pandora's hope: essays on the reality of science studies*. Harvard university press.

Lazarsfeld, Paul F, Robert K Merton, et al. (1954). "Friendship as a social process: A substantive and methodological analysis". In: *Freedom and control in modern society* 18.1, pp. 18–66.

Lee, John D and Katrina A See (2004). "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1, pp. 50–80.

Lee, Michael D and Eric-Jan Wagenmakers (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Lee, Min Kyung and Jodi Forlizzi (2009). "Designing adaptive robotic services". In: *Proc. of IASDR'09*.

Lee, Min Kyung et al. (2010). "Gracefully Mitigating Breakdowns in Robotic Services". In: *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*. HRI '10. Osaka, Japan: IEEE Press, pp. 203–210. ISBN: 978-1-4244-4893-7. URL: http://dl.acm.org.proxy.library.cmu.edu/citation.cfm?id=1734454.1734544.

Lee, Min Kyung et al. (2012). "Personalization in HRI: A longitudinal field experiment". In: *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 319–326.

Li, Jamy (May 2015). "The Benefit of Being Physically Present". In: *Int. J. Hum.-Comput. Stud.* 77.C, pp. 23–37. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2015.01.001. URL: http://dx.doi.org/10.1016/j.ijhcs.2015.01.001.

Lieberman, Henry and Ted Selker (2003). "Agents for the user interface". In: *Handbook of Agent Technology*. Publisher: MIT Press, pp. 1–21.

Lim, Kai H, Lawrence M Ward, and Izak Benbasat (1997). "An empirical study of computer system learning: Comparison of co-discovery and self-discovery methods". In: *Information Systems Research* 8.3, pp. 254–272.

Lindley, Joseph, Adrian Gradinar, and Paul Coulton (2020). "Ghosts in the Smart Home". In: *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*. DIS' 20 Companion. Eindhoven, Netherlands: Association for Computing Machinery, pp. 465468. ISBN: 9781450379878. DOI: 10.1145/3393914.3395841. URL: https://doi.org/10.1145/3393914.3395841.

Litoiu, Alexandru et al. (2015). "Evidence That Robots Trigger a Cheating Detector in Humans". In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '15. Portland, Oregon, USA: ACM, pp. 165–172. ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696456. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/2696454.2696456.

Luger, Ewa and Abigail Sellen (2016). ""Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, pp. 5286–5297. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858288. URL: http://doi.acm.org/10.1145/2858036.2858288.

Lupton, Deborah (2020). *The story completion method and more-than-human theory: Finding and using health information*. Thousand Oaks, CA: SAGE Publications Ltd.

Luria, Michal, Guy Hoffman, and Oren Zuckerman (May 2017). "Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control". en. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver Colorado USA: ACM, pp. 580–628. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025786. URL: https://dl.acm.org/doi/10.1145/3025453.3025786 (visited on 07/22/2020).

Luria, Michal et al. (2019). "Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents". In: *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 633–644.

Luria, Michal et al. (Apr. 2020). "Social Boundaries for Personal Agents in the Interpersonal Space of the Home". en. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, pp. 1–12. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376311. URL: https://dl.acm.org/doi/10.1145/3313831.3376311.

Löwgren, Jonas (Jan. 2013). "Annotated portfolios and other forms of intermediate-level knowledge". In: *Interactions* 20.1, pp. 30–34. ISSN: 1072-5520. DOI: 10.1145/2405716.2405725. URL: http://doi.org/10.1145/2405716.2405725.

Ma, Lanssie Mingyue et al. (2018). "Human-Robot Teaming: Concepts and Components for Design". en. In: *Field and Service Robotics*. Ed. by Marco Hutter and Roland

Siegwart. Vol. 5. Series Title: Springer Proceedings in Advanced Robotics. Cham: Springer International Publishing, pp. 649–663. ISBN: 978-3-319-67360-8 978-3-319-67361-5. DOI: 10.1007/978-3-319-67361-5_42. URL: http://link.springer.com/10.1007/978-3-319-67361-5_42 (visited on 06/11/2020).

Maes, Pattie (July 1994). "Agents that reduce work and information overload". In: *Communications of the ACM* 37.7, pp. 30–40. ISSN: 00010782. DOI: 10.1145/176789.176792. URL: http://portal.acm.org/citation.cfm?doid=176789.176792 (visited on 06/18/2020).

Marquardt, Nicolai et al. (2009). "Situated messages for asynchronous human-robot interaction". In: *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 301–302.

Martin, Alan et al. (2005). "Maintaining the identity of dynamically embodied agents". In: *International Workshop on Intelligent Virtual Agents*. Springer, pp. 454–465.

Martin, David L., Adam J. Cheyer, and Douglas B. Moran (1999). "The open agent architecture: A framework for building distributed software systems". In: *Applied Artificial Intelligence* 13.1-2. Publisher: Taylor & Francis, pp. 91–128.

McCaffrey, Chloe et al. (2021). "Can robots be used to encourage social distancing?" In: *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 475–478.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks". In: *Annual Review of Sociology* 27.1, pp. 415–444.

Meet Vector (n.d.). en. URL: https://www.digitaldreamlabs.com/pages/meet-vector (visited on 01/01/2021).

Mittal, Banwari and Walfried M Lassar (1996). "The role of personalization in service encounters". In: *Journal of Retailing* 72.1, pp. 95–109.

Moharana, Sanika et al. (2019). "Robots for Joy, Robots for Sorrow: Community Based Robot Design for Dementia Caregivers". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 458–467.

Morales, Cecilia G. et al. (June 2019). "Interaction Needs and Opportunities for Failing Robots". en. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*. San Diego CA USA: ACM, pp. 659–670. ISBN: 978-1-4503-5850-7. DOI: 10.1145/3322276.3322345. URL: https://dl.acm.org/doi/10.1145/3322276.3322345 (visited on 09/21/2020).

Muir, Bonita Marlene (1989). *Operators' Trust in and Use of Automatic Controllers in a Supervisory Process Control Task*. en. Google-Books-ID: lESrjgEACAAJ. University of Toronto. ISBN: 978-0-315-51014-2.

Muir, Bonnie Marlene (2002). "Operators' trust in and use of automatic controllers in a supervisory process control task." In.

Mutlu, B., J. Forlizzi, and J. Hodgins (2006). "A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior". In: *2006 6th IEEE-RAS International Conference on Humanoid Robots*, pp. 518–523. DOI: 10.1109/ICHR.2006.321322.

Mutlu, Bilge et al. (2012). "Conversational gaze mechanisms for humanlike robots". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1.2, pp. 1–33.

Nass, Clifford, BJ Fogg, and Youngme Moon (1996). "Can computers be teammates?" In: *International Journal of Human-Computer Studies* 45.6, pp. 669–678.

Nass, Clifford, Jonathan Steuer, and Ellen R Tauber (1994). "Computers are social actors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 72–78.

Nass, Clifford et al. (1995). "Can computer personalities be human personalities?" In: *Conference companion on Human factors in computing systems*. ACM, pp. 228–229.

Nelder, JA458743 (1977). "A reformulation of linear models". In: *Journal of the Royal Statistical Society: Series A (General)* 140.1, pp. 48–63.

Odom, William et al. (2012). "A fieldwork of the future with user enactments". In: *Proceedings of the Designing Interactive Systems Conference*. ACM, pp. 338–347.

Odom, William et al. (2014). "Unpacking the thinking and making behind a user enactments project". In: *Proceedings of the 2014 conference on Designing interactive systems*, pp. 513–522.

Ogawa, Kohei and Tetsuo Ono (2008). "ITACO: Effects to interactions by relationships between humans and artifacts". In: *International Workshop on Intelligent Virtual Agents*. Springer, pp. 296–307.

Oistad, Benjamin C. et al. (2016). "Colleague or Tool? Interactivity Increases Positive Perceptions of and Willingness to Interact with a Robotic Co-worker". In: *International Conference on Social Robotics*. Springer, pp. 774–785.

Onnasch, Linda and Clara Laudine Hildebrandt (2021). "Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction". In: *ACM Transactions on Human-Robot Interaction (THRI)* 11.1, pp. 1–24.

Paiva, Ana (2017). *The LIREC Project (2008-2012)*. URL: https://ana-paiva.com/home/about/projects/previous-projects/.

Pan, Rui, Henry Lo, and Carman Neustaedter (2017). "Collaboration, Awareness, and Communication in Real-Life Escape Rooms". In: *Proceedings of the 2017 Conference on Designing Interactive Systems*. DIS '17. Edinburgh, United Kingdom: ACM, pp. 1353–1364. ISBN: 978-1-4503-4922-2. DOI: 10.1145/3064663.3064767. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/3064663.3064767.

Pandey, Amit Kumar and Rodolphe Gelin (2018). "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind". In: *IEEE Robotics & Automation Magazine* 25.3, pp. 40–48.

Patterson, HD (1975). "Maximum likelihood estimation of components of variance". In: *Proceeding Eight International Biometric Conference, 1975*. Biometric Soc.

Perlmutter, Leah et al. (2016). "Automatic Adaptation of Online Language Lessons for Robot Tutoring". In: *International Conference on Social Robotics*. Springer, pp. 660–670.

Powers, Aaron et al. (2005). "Eliciting information from people with a gendered humanoid robot". In: *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, pp. 158–163.

Powers, Aaron et al. (2007). "Comparing a Computer Agent with a Humanoid Robot". In: *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*. HRI '07. Arlington, Virginia, USA: ACM, pp. 145–152. ISBN: 978-1-59593-617-2. DOI: 10.1145/1228716.1228736. URL: http://doi.acm.org/10.1145/1228716.1228736.

Purington, Amanda et al. (2017). ""Alexa is My New BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, pp. 28532859. ISBN: 9781450346566. URL: https://doi.org/10.1145/3027063.3053246.

Rapp, Adam A et al. (2017). “The role of the sales-service interface and ambidexterity in the evolving organization: A multilevel research agenda”. In: *Journal of Service Research* 20.1, pp. 59–75.

Reig, Samantha et al. (2018). “A field study of pedestrians and autonomous vehicles”. In: *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pp. 198–209.

Reig, Samantha et al. (Mar. 2020). “Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot”. In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’20. Cambridge, United Kingdom: Association for Computing Machinery, pp. 289–297. ISBN: 978-1-4503-6746-2. DOI: 10.1145/3319502.3374795. URL: https://doi.org/10.1145/3319502.3374795 (visited on 05/06/2020).

Reig, Samantha et al. (2021a). “Flailing, Hailing, Prevailing: Perceptions of Multi-Robot Failure Recovery Strategies”. In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 158–167.

Reig, Samantha et al. (2021b). “Social Robots in Service Contexts: Exploring the Rewards and Risks of Personalization and Re-embodiment”. In: *Designing Interactive Systems Conference 2021*, pp. 1390–1402.

Ringberg, Torsten, Gaby Odekerken-Schröder, and Glenn L Christensen (2007). “A cultural models approach to service recovery”. In: *Journal of Marketing* 71.3, pp. 194–214.

Robinette, Paul, Ayanna M Howard, and Alan R Wagner (2017). “Effect of robot performance on human–robot trust in time-critical situations”. In: *IEEE Transactions on Human-Machine Systems* 47.4, pp. 425–436.

Salomons, Nicole et al. (Feb. 2018). “Humans Conform to Robots: Disambiguating Trust, Truth, and Conformity”. en. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. Chicago IL USA: ACM, pp. 187–195. ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171282. URL: https://dl.acm.org/doi/10.1145/3171221.3171282 (visited on 06/23/2020).

Salvini, Pericle et al. (2010). “How safe are service robots in urban environments? Bullying a robot”. In: *19th international symposium in robot and human interactive communication*. IEEE, pp. 1–7.

Scheutz, Matthias, Scott A. DeLoach, and Julie A. Adams (2017). “A framework for developing and using shared mental models in human-agent teams”. In: *Journal of Cognitive Engineering and Decision Making* 11.3. Publisher: Sage Publications Sage CA: Los Angeles, CA, pp. 203–224.

Scheutz, Matthias and Bertram F Malle (2018). “Moral robots.” In.

Schiffhauer, Birte et al. (2016). “Let the User Decide! User Preferences Regarding Functions, Apps, and Interfaces of a Smart Home and a Service Robot”. In: *International Conference on Social Robotics*. Springer, pp. 971–981.

Sciuto, Alex et al. (2018). “" Hey Alexa, What’s Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage”. In: *Proceedings of the 2018 designing interactive systems conference*, pp. 857–868.

Seering, Joseph et al. (2020). "It Takes a Village: Integrating an Adaptive Chatbot into an Online Gaming Community". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

Segura, Elena Márquez et al. (2012). "How Do You Like Me in This: User Embodiment Preferences for Companion Agents". en. In: *Intelligent Virtual Agents*. Ed. by David Hutchison et al. Vol. 7502. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 112–125. ISBN: 978-3-642-33196-1 978-3-642-33197-8. DOI: 10.1007/978-3-642-33197-8_12. URL: http://link.springer.com/10.1007/978-3-642-33197-8_12 (visited on 01/08/2019).

Sellen, Abigail et al. (2006). "HomeNote: supporting situated messaging in the home". en. In: *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work - CSCW '06*. Banff, Alberta, Canada: ACM Press, p. 383. ISBN: 978-1-59593-249-5. DOI: 10.1145/1180875.1180933. URL: http://portal.acm.org/citation.cfm?doid=1180875.1180933 (visited on 01/05/2021).

Sembroski, Catherine E., Marlena R. Fraune, and Selma Šabanović (2017). "He Said, She Said, It Said: Effects of Robot Group Membership and Human Authority on People's Willingness to Follow Their Instructions". In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 56–61.

Shakeri, Hanieh et al. (2017). "Escaping Together: The Design and Evaluation of a Distributed Real-Life Escape Room". In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '17. Amsterdam, The Netherlands: ACM, pp. 115–128. ISBN: 978-1-4503-4898-0. DOI: 10.1145/3116595.3116601. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/3116595.3116601.

Short, Elaine et al. (2010). "No Fair!!: An Interaction with a Cheating Robot". In: *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*. HRI '10. Osaka, Japan: IEEE Press, pp. 219–226. ISBN: 978-1-4244-4893-7. URL: http://dl.acm.org.proxy.library.cmu.edu/citation.cfm?id=1734454.1734546.

Singer, Peter Warren (2009). *Wired for war: The robotics revolution and conflict in the 21st century*. Penguin.

Singhal, Samarth et al. (2018). "Time-Turner: Designing for Reflection and Remembrance of Moments in the Home". en. In: ACM Press, pp. 1–14. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173753. URL: http://dl.acm.org/citation.cfm?doid=3173574.3173753 (visited on 06/20/2018).

Sorrentino, Mike (2018). *Sony's adorable and expensive AIBO puppy is now on sale in the US*. URL: https://www.cnet.com/news/sonys-adorable-and-expensive-aibo-puppy-is-now-on-sale-in-the-us/.

SRI (n.d.). *Cognitive assistant that learns and organizes*. URL: https://www.sri.com/hoi/artificial-intelligence-calo/.

Stankovic, John A. (Feb. 2014). "Research Directions for the Internet of Things". In: *IEEE Internet of Things Journal* 1.1. Conference Name: IEEE Internet of Things Journal, pp. 3–9. ISSN: 2327-4662. DOI: 10.1109/JIOT.2014.2312291.

Stoll, Brett et al. (2018). "Wait, Can You Move the Robot?: Examining Telepresence Robot Use in Collaborative Teams". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. Chicago, IL, USA: ACM, pp. 14–22. ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171243. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/3171221.3171243.

Stolterman, Erik (2008). “The nature of design practice and implications for interaction design research”. In: *International Journal of Design* 2.1. Publisher: Chinese Institute of Design.

Streitz, Norbert (2019). “Beyond smart-onlycities: redefining the smart-everythingparadigm”. In: *Journal of Ambient Intelligence and Humanized Computing* 10.2. Publisher: Springer, pp. 791–812.

Streitz, Norbert et al. (Jan. 2019). “Grand challenges for ambient intelligence and implications for design contexts and smart societies”. en. In: *Journal of Ambient Intelligence and Smart Environments* 11.1, pp. 87–107. ISSN: 18761372, 18761364. DOI: 10.3233/AIS-180507. URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AIS-180507 (visited on 07/29/2020).

Streitz, Norbert A. et al. (2005). “Designing smart artifacts for smart environments”. In: *Computer* 38.3. Publisher: IEEE, pp. 41–49.

Strohkorb Sebo, Sarah et al. (2018). “The ripple effects of vulnerability: The effects of a robot’s vulnerable behavior on trust in human-robot teams”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 178–186.

Stroup, Walter W (2016). *Generalized linear mixed models: modern concepts, methods and applications*. CRC press.

Stuck, Rachel E. et al. (2017). “Understanding Attitudes of Adults Aging with Mobility Impairments Toward Telepresence Robots”. In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’17. Vienna, Austria: ACM, pp. 293–294. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3038351. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/3029798.3038351.

Suguitan, Michael and Guy Hoffman (2019). “Blossom: A Handcrafted Open-Source Robot”. In: *ACM Transactions on Human-Robot Interaction (THRI)* 8.1, p. 2.

Sung, JaYoung, Rebecca E Grinter, and Henrik I Christensen (2009). “Pimp my roomba: Designing for personalization”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 193–196.

Sycara, Katia P (1998). “Multiagent Systems”. en. In: p. 15.

Syrdal, Dag Sverre et al. (2009). “The boy-robot should bark!-childrens impressions of agent migration into diverse embodiments”. In: *Proceedings: New Frontiers of Human-Robot Interaction, a symposium at AISB*.

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

Takayama, L., W. Ju, and C. Nass (2008). “Beyond dirty, dangerous and dull: What everyday people think robots should do”. In: *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 25–32. DOI: 10.1145/1349822.1349827.

Tan, Xiang Zhi et al. (2019). “From One to Another: How Robot-Robot Interaction Affects Users’ Perceptions Following a Transition Between Robots”. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 114–122.

Taub, Eric A (2006). *For Sony’s Robotic Aibo, it’s the last year of the dog*. URL: https://www.nytimes.com/2006/01/30/technology/for-sonys-robotic-aibo-its-the-last-year-of-the-dog.html.

Tay, Benedict, Younbo Jung, and Taezoon Park (2014). “When stereotypes meet robots: the double-edge sword of robot gender and personality in humanrobot interaction”. In: *Computers in Human Behavior* 38. Publisher: Elsevier, pp. 75–84.

Tejwani, Ravi et al. (July 2020). “Migratable AI”. In: *arXiv:2007.05801 [cs]*. arXiv: 2007.05801. URL: http://arxiv.org/abs/2007.05801 (visited on 07/23/2020).

Tellex, Stefanie et al. (2014). “Asking for help using inverse semantics”. In: Publisher: Robotics: Science and Systems Foundation.

Thomas, Frank, Ollie Johnston, and Frank Thomas (1995). *The illusion of life: Disney animation*. Hyperion New York.

Triandis, Harry C. (2001). “Individualism-collectivism and personality”. In: *Journal of personality* 69.6. Publisher: Wiley Online Library, pp. 907–924.

Troiano, Giovanni Maria, Matthew Wood, and Casper Harteveld (2020). “"And This, Kids, Is How I Met Your Mother": Consumerist, Mundane, and Uncanny Futures with Sex Robots”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, pp. 117. ISBN: 9781450367080. DOI: 10.1145/3313831.3376598. URL: https://doi.org/10.1145/3313831.3376598.

Truong, Khai N., Gillian R. Hayes, and Gregory D. Abowd (2006). “Storyboarding: an empirical determination of best practices and effective guidelines”. In: *Proceedings of the 6th conference on Designing Interactive systems*, pp. 12–21.

Tsujimoto, Megumi, N Munekat, and Tetsuo Ono (2013). “Evaluating how the humans impression formation of robots is effected by the relation between the robots”. In: *Proceedings of the 1st International Conference on Human-Agent Interaction. Tsukuba, Japan*.

Tulli, Silvia et al. (2019). “Great Expectations & Aborted Business Initiatives: The Paradox of Social Robot Between Research and Industry.” In: *BNAIC/BENELEARN*.

Ullman, Daniel and Bertram F Malle (2019). “Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust”. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 618–619.

Vázquez, Marynel et al. (2011). “A Deceptive Robot Referee in a Multiplayer Gaming Environment”. In: *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 204 –211.

Vázquez, Marynel et al. (2017). “Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze”. In: *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, pp. 42–52.

Verbeek, Peter-Paul (2006). “Materializing morality: Design ethics and technological mediation”. In: *Science, Technology, & Human Values* 31.3, pp. 361–380.

— (Apr. 2015). “Beyond Interaction: A Short Introduction to Mediation Theory”. In: *Interactions* 22.3, pp. 26–31. ISSN: 1072-5520. DOI: 10.1145/2751314. URL: https://doi.org/10.1145/2751314.

Wang, Bingcheng and Pei-Luen Patrick Rau (2018). “Influence of Embodiment and Substrate of Social Robots on Users Decision-Making and Attitude”. In: *International Journal of Social Robotics*, pp. 1–11.

Watson, Ash and Deborah Lupton (2020). “Tactics, affects and agencies in digital privacy narratives: a story completion study”. In: *Online Information Review* 45.1, pp. 138–156. DOI: 10.1108/OIR-05-2020-0174.

Weiss, Astrid et al. (2010). “Robots asking for directions - The willingness of passers-by to support robots”. In: *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 23–30.

Wen, Ruchen et al. (2021). “Comparing strategies for robot communication of role-grounded moral norms”. In: *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 323–327.

Williams, Kristin et al. (2020). “The Upcycled Home: Removing Barriers to Lightweight Modification of the Home’s Everyday Objects”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

Williams, Tom et al. (2021). “Deconstructed Trustee Theory: Disentangling Trust in Body and Identity in Multi-Robot Distributed Systems”. In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 262–271.

Winkle, Katie et al. (2022). “Norm-breaking responses to sexist abuse: A cross-cultural human robot interaction study”. In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 120–129.

Wong, Richmond Y and Deirdre K Mulligan (2016). “When a product is still fictional: anticipating and speculating futures through concept videos”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 121–133.

Wood, Matthew, Gavin Wood, and Madeline Balaam (2017). “"They’re Just Tixel Pits, Man": Disputing the ’Reality’ of Virtual Reality Pornography through the Story Completion Method”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, pp. 54395451. ISBN: 9781450346559. URL: https://doi.org/10.1145/3025453.3025762.

Wooldridge, Michael and Nicholas R. Jennings (1995). “Intelligent agents: Theory and practice”. In: *The knowledge engineering review* 10.2. Publisher: Cambridge University Press, pp. 115–152.

Wu, Xian et al. (2017). “An Evaluation of a Telepresence Robot: User Testing Among Older Adults with Mobility Impairment”. In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’17. Vienna, Austria: ACM, pp. 325–326. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3038324. URL: http://doi.acm.org.proxy.library.cmu.edu/10.1145/3029798.3038324.

Young, James E and Ehud Sharlin (2006). “Sharing spaces with robots An integrated environment for human-robot interaction”. In.

Zimmerman, John and Jodi Forlizzi (Mar. 2017). “Speed Dating: Providing a Menu of Possible Futures”. en. In: *She Ji: The Journal of Design, Economics, and Innovation* 3.1, pp. 30–50. ISSN: 2405-8726. DOI: 10.1016/j.sheji.2017.08.003. URL: http://www.sciencedirect.com/science/article/pii/S2405872616300661 (visited on 05/06/2020).

Zimmerman, John, Jodi Forlizzi, and Shelley Evenson (2007). “Research through design as a method for interaction design research in HCI”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 493–502.