AN ABSTRACT OF THE DISSERTATION OF

Matthew Rueben for the degree of Doctor of Philosophy in Robotics presented on November 20, 2018.

Title: Privacy-Sensitive Robotics

Abstract approved: _____

William D. Smart

This dissertation focuses on personal privacy in human-robot interaction, which we call "privacy-sensitive robotics." Our understanding of "privacy" is very broad, including not just information privacy but also physical, psychological, and social privacy. We begin by surveying the scholarly literature on privacy and talking about why it applies to interactions with robots. We then make five contributions to help launch privacy-sensitive robotics as an emerging area of research—one from a literature review, three from empirical studies, and one about the future of privacy-sensitive robotics research:

1. We begin by presenting the current state of the art in privacy protection technologies (whether or not they were designed as such) from the literature.

2. Our first study found differences in usability and user preference between three different interfaces for specifying user privacy preferences to a robot.

3. Our next study showed how the contextual "framing" of an action affects whether people see it as a privacy violation.

4. Our third and final study documents the process of forming beliefs about the robot's sensing capabilities and identifies some key aspects of this process for further study.

5. Finally, we give a set of recommendations for developing privacy-sensitive robotics as a research area.

These five contributions are linked by the goal of privacy-sensitive robotics research: to enable a future in which robotics technology upholds and respects our privacy. We close with a call to action for potential privacy-sensitive robotics researchers.

Privacy-Sensitive Robotics

by

Matthew Rueben

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented November 20, 2018
Commencement June 2019

Doctor of Philosophy dissertation of Matthew Rueben presented on November 20, 2018.

APPROVED:

Major Professor, representing Robotics

Head of the School of Mechanical, Industrial, and Manufacturing Engineering

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Matthew Rueben, Author

ACKNOWLEDGEMENTS

I would like to thank ...

Bill and Cindy for all your guidance and input these five-and-a-half years. I would be honored to be your colleague someday.

Frank Bernieri for all the teaching, encouragement, and long conversations that made social psychology my second language. I appreciated connecting with you as a fellow academic at heart.

My fellow HRI privacy researchers, including Ross Sowell, Maya Cakmak, Christoph Lutz, Meg Tonkin and Jon Vitale, and Eduard Fosch Villaronga.

Also, my fellow privacy researchers from the We Robot community, including Ryan Calo, Margot Kaminski, Woody Hartzog, Christopher Millard, Hideyuki ("Yuki") Matsumi, and Ashkan Soltani.

Several of my co-authors on privacy papers: Margaret Krupp, Alex Hubers, Alex Aroyo, and Johannes Schmölz.

People who have had a big effect on my path through grad school, including Jonathan Hurst for helping me decide (rightly, I think) to do a Ph.D. instead of a Master's, as well as Ross Sowell for painting an attractive picture of life as a teaching professor.

Grad students like Kory Kraft and professors like Ken Funk, Gary Ferngren, and Andy Karplus who model how to be a Christian in academia . . . and encourage me to go and do likewise.

People who show that successful HRI professors can also be kind and warm—people like Cindy Bethel, Selma Šabanović, and Friederike Eyssel.

Some of my peers whom I now also consider my friends: Marlena Fraune, Maike Paetzel, Leah Perlmutter, and Eduard Fosch Villaronga.

Also, all the other kind people at conferences who have invited me along to lunches and dinners and made me feel included when I didn't know anybody.

Friends and mentors from FIRST Robotics Team 955 for getting me hooked on robotics—especially Will Rottenkolber for being such an influential mentor to me.

Rob Holman and Dan Cox at the OSU Wave Lab for making me learn MATLAB by myself.

My mentors at each of my internships: Matt Trippel at Vestas, Gary Hicks at CH2M HILL, and Maarten Bos at Disney Research.

Duy Nguyen for being the only other "social" person in the lab with me, and for being my friend.

Yawei Zhang for his generous heart, for making me laugh, and for being my friend.

Jeff Klow for being a higher-level wizard than I am.

My Corner House family, especially Peter and Hilary for hosting me during the final stretch of dissertation writing.

"The gang:" Jenna Proctor, Peter Killgore, and Megan Watts.

My Christian Grad Fellowship family for being the way God's family should be.

James Roberts, my good friend, for your loyalty and support.

My Mom and Dad, who have really "shared" this journey with me. Thanks for cooking me dinner and listening to me talk for hours and hours.

Most importantly: Jesus Christ, my biggest supporter, by whose plan and power all these other people were there for me. Having made all good things, he is the ultimate authority on making all things (including robots) good.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

Privacy-Sensitive Robotics

1 Introduction

When humans look at the world we go beyond mere facts to *values*—i.e., the importance, worth, or desirability of objects and states of affairs. We can value concrete or mundane things as well as abstract or lofty ones: a drink of water, my next-door neighbor, justice, my favorite hat, the people of Oregon, a walk in the woods, regular meals with my family, and friendship[1]. As we develop robotics technologies, they will interact more and more with the things we value. To the extent that the robots we have made so far have been for *industrial* tasks like moving heavy things or performing rapid, precise assembly tasks, we are used to designing robots to optimize values like speed, accuracy, and cost. As we turn towards making robots that operate outside of industrial settings, however, and especially as they become a part of our *social* lives, they will begin to interact with other things we value: personal property, privacy, healthy relationships, courtesy, calmness, and human autonomy and identity[2]. These other, more social values can be very important, but also easy to overlook or, worse, intentionally ignore at users' expense. Neglecting some of these values has caused such big problems in the past that many companies have instituted formal processes for protecting them; examples include safety [67], usability [248], and accessibility [3]. It would

[1]This explanation of human values is inspired by the one in Section 2.1 of "Value Sensitive Design and Information Systems" [80].

[2]This list is adapted from the one in Table 1, ibid.

not be wise to always wait until significant damage has been done before we learn to consider *all* the important values that robotics technologies can impact.

One systematic way to consider what people value when we create or evaluate technologies is through Value Sensitive Design [80]. Value Sensitive Design is a three-part methodology: (1) investigate the values involved ("conceptual investigation"), (2) investigate how people experience and prioritize those values in real life ("empirical investigation"), (3) investigate how the technology in question protects or damages the things we value ("technical investigation"). The work presented in this dissertation represents the start of a new effort to bring much-needed consideration to a value that has been neglected so far by the robotics community: privacy.

Privacy can be defined as the effective setting of boundaries between oneself and other people [10]. These boundaries can regulate personal information, personal space, territory, social interaction, relationships, thoughts, feelings, opinions, and decisions [196]. All humans need privacy, although different cultures seek it in different ways [11]. Privacy is a crucial requirement for relationships to flourish [110] and for individuals to mature and have freedom. As robots enter human-occupied spaces, they will pose new threats to the privacy of the people around them. Robots are capable of violating human privacy: they can collect and share information, move through personal spaces and territories, and interact with people socially [168]. Thus, robots and the companies that design them will need to earn people's trust regarding privacy if they are going to be accepted.

Research has not given us the knowledge we need to do this. Privacy scholars

have formulated several theories of how privacy works between humans [10, 172], but we do not know how they extend to interactions with robots. Plus, privacy has several distinct dimensions [196] and is often the subject of legal debates—privacy is a complicated value. The field of human-robot interaction (HRI) has barely begun to think about privacy—it has neither drawn significantly from privacy scholarship nor designed practical systems to comprehensively protect user privacy.

In this dissertation we present work to help launch a new surge of research on privacy in HRI—we call this sort of work "privacy-sensitive robotics" research. We present findings and recommendations from our delvings into privacy scholarship as well as from three empirical studies of human-robot interactions.

The following chapter (Chapter 2) presents our understanding of privacy and why it matters for human-robot interactions from a detailed review of the privacy literature. We then present five contributions to "privacy-sensitive robotics" research—the first is a survey of existing technologies, the next three are findings from empirical studies, and the last is a set of recommendations:

1. **Survey of Existing Technologies.** To protect users' privacy we will need technologies that enforce *constraints* on robot perception, navigation, and manipulation. We survey the literature and present some existing technologies that have already been used as constraints on robots, as well as many more that could potentially be used for this purpose in the future (Chapter 3).

2. **Study of Interfaces for Specifying Privacy Preferences.** For our first empirical contribution we focused on visual privacy by evaluating "physical

marker" interfaces for specifying private objects. We present findings from a controlled experiment in a real office setting using prototypes of the interfaces with a PR2 robot (Chapter 4).

3. **Study of Contextual "Framing" Effects on Privacy Judgments.** For our second empirical contribution we present findings from four online surveys on the effects of contextual framing on privacy judgments about a telepresence robot. We used realistic, animated videos of a telepresence robot in a home setting (Chapter 5).

4. **Study of Mental Model Formation over Time in the Wild.** For our third and final empirical contribution we interviewed participants throughout a six week study of multiple interactions with a novel robot in a natural setting. We present findings about how each user formed their mental model of how the robot works (Chapter 6).

5. **Recommendations for the Future of Privacy-Sensitive Robotics.** We present a roadmap for this new area of research, as well as the various topics to study and the collaborations needed to study them, for our final contribution (Chapter 7).

These five contributions are linked by the goal of privacy-sensitive robotics research: to enable a future in which robotics technology upholds and respects our privacy. We close with a short summary of our findings and recommendations, plus a call to action for privacy-sensitive robotics research (Chapter 8).

Note that although we are not design experts and do not go through an entire design process for a particular robotics application, we do work relevant to all three parts of Value Sensitive Design [80] in this dissertation. We do conceptual investigation in the next chapter (Chapter 2), technical investigation in Chapters 3 and 4, and empirical investigation in Chapters 4, 5, and 6. We intend for this dissertation to begin showing how to do each of these three types of investigation for privacy concerns about the use of robots.

2 Privacy in Human-Robot Interaction: Background and Literature Review

This chapter introduces the emerging research area we call "privacy-sensitive robotics." We begin by discussing different definitions of privacy and surveying the privacy literature across several disciplines. We then turn to the question of how privacy is implicated by human-robot interactions. After introducing "privacy-sensitive robotics" as a new area of research, we outline the rest of this dissertation with a preview of our five contributions.

2.1 What is Privacy?

There is a lot of literature on privacy. The conversation spans many disciplines and is difficult to summarize concisely. The following several sections (this one plus Sections 2.2, 2.3, and 2.4) present our summary of privacy scholarship beginning with abstract ideas and proceeding towards concrete applications in the real world. This first section is about the *definition* of privacy, over which much ink has been spilled.

2.1.1 Views on the Definition of Privacy

Perhaps the simplest definition of privacy is that of Judge Thomas Cooley: the right "to be let alone" [251]. It becomes clear that privacy is difficult to define without generalizing or inviting criticism. What follows is an aggregation of many definitions and descriptions of what privacy is. We will then give the definitions *we* used for our work in the form of theories (Section 2.1.2), a taxonomy (Section 2.1.3), and a discussion of how the word is used in everyday conversation (Section 2.1.4).

We recommend the Stanford Encyclopedia of Philosophy article on privacy by Judith DeCew as a comprehensive guide to the definition of privacy [58], especially in law and philosophy. Most of the references in this section we owe to the bibliography from that article. As we explore the various *definitions* of privacy in the literature, we will follow her threefold division: informational privacy, constitutional privacy (i.e., in making intimate decisions about oneself), and privacy in terms of *access* to the self, both physical and mental.

2.1.1.1 Informational Privacy.

Informational privacy refers to privacy concerns about personal information. This was the definition of privacy held by Warren and Brandeis [251] in their famous article. A number of other authors have presented informational definitions of privacy, although whether they believed this to be the *only* aspect of privacy is not in view here. Prosser [185] divided (informational) privacy into four parts. His for-

mulation continues to be referenced today. Briefly, Prosser divides (informational) privacy into intrusion, public disclosure, false light, and appropriation. These mean the following. First, intrusion into one's private affairs includes trespassing, search, and remote intrusion such as wire tapping. Second is public disclosure of private facts. Third is publicly portraying the victim in a false light, e.g., by misattributing to the victim a statement or opinion. Fourth is appropriation, or pretending to be the victim for one's own advantage. Solove [223] has constructed a taxonomy of privacy concepts based on Prosser's formulation. It is shown in Figure 2.1 as a general overview of informational privacy concerns.

A number of other authors offer informational definitions of privacy that seem to gel with Prosser's. Fried [78] defines privacy as control over knowledge about oneself. Parent [180] defines privacy as the condition of others not possessing undocumented information about oneself. Parent also follows, to a certain extent, in the footsteps of so-called "privacy reductionists" like Thomson [241] in that he dismisses other, non-informational aspects of privacy as being covered by other rights. For example, Parent considers constitutional privacy concerns about the freedom to make important personal decisions to be a matter simply of liberty. Privacy is not a branch of liberty, he argues, because they could be at odds; for example, when someone freely gives up their privacy they are simultaneously exercising their liberty! Parent argues against Thomson, however, by keeping the right to privacy as a distinctive idea; Parent believes that Thomson had to invent some far-fetched human rights in order to maintain her view. "Privacy reductionism" is discussed further in Section 2.2.1.

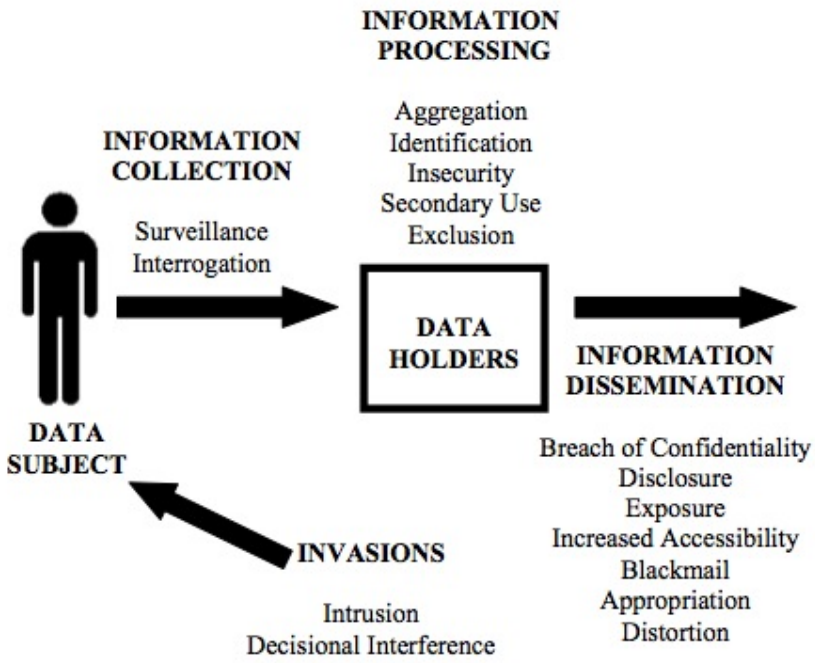

Figure 2.1: Daniel Solove's visual "model" of his taxonomy of (informational) privacy [223].

Moore [164] defines privacy as, "control over access to oneself and information about oneself." This is a "control-based" definition of privacy, in which it doesn't matter *per se* whether somebody accesses you or your information, but rather whether you can control that access. Control-based definitions account for situations in which someone invites others into his close company, or willingly gives out personal information. These actions would violate privacy if privacy is the state of being let alone, or of having all your personal information kept to yourself. But authors holding to control-based definitions of privacy maintain that the person in question is still in control, so there's no violation; this especially makes sense in the legal context. Moore [163] defines this control in terms of the inner spheres of personal information, about the "core self." He limits his definition to the "core self" to intentionally limit the First Amendment[1] freedom of speech in some cases. Moore argues that sex offenders *should* be forced to reveal their criminal history to their neighbors, and that politicians, having chosen a very public profession, cannot complain when the public learns about some of their more outward characteristics. On the other hand, Moore says, it is suspicious whenever the press claims that the public has a "right to know the truth" about some controversial but very private incident. He holds that privacy may win out in such a situation.

Austin [17] offers a more nuanced definition of privacy: freedom from "public gaze." She argues that this updated definition deals with the problem of new tech-

[1]The First Amendment to the U.S. Constitution reads in its entirety, "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances." [2]

nologies to which older definitions of privacy do not object. In particular, Austin is concerned about cases wherein people know they are under surveillance, about the collection of non-intimate but personal information (e.g., in data mining), and about the collection of personal information in public. She claims that other, older definitions of privacy do not agree with our intuition that these technologies (could) invade our privacy by denying us our freedom from "public gaze."

2.1.1.2 Constitutional Privacy and Autonomy.

A second legal interpretation of privacy emerged in 1965 with the *Griswold v. Connecticut* decision [255]. This was named the "constitutional right to privacy," defined rather vaguely as, "a right protecting one's individual interest in independence in making certain important and personal decisions about one's family, life and lifestyle" [58]. Note that the protection is against governmental, not private, action. This has been used to overturn laws against certain sexual acts and against abortion. In law, this represents a general shift from property-based definitions of privacy to definitions based more on personal liberty. The constitutional right to privacy is mentioned here as a nod to its importance in the conversation about defining privacy; since it applies very little to technology and robotics, it is largely ignored outside of this section and the section on privacy in U.S. law (Section 2.2.2).

2.1.1.3 Access Privacy: Intimacy, Solitude, and Space.

Daniel Solove wrote that a society without privacy is a "suffocating society" [223]. This section presents the broadest conception of privacy, which concerns the human need for being away from others for the sake of both the self and one's intimate relationships. Along these lines, Alan Westin lists four different states of privacy: solitude, anonymity, intimacy (i.e., being alone *with* someone), and reserve (i.e., keeping to oneself) [255]. Similarly, Leino-Kilpi et al. [144] divide privacy into (1) *physical privacy*, over personal space or territory; (2) *psychological privacy*, over thoughts and values; (3) *social privacy*, over interactions with others and influence from them; and (4) *informational privacy*, over personal information. This section reviews the definitions of privacy from the first three of those categories—physical, psychological, and social privacy—and from all four of Westin's private states. These conceptions of privacy are of special interest to anthropologists, psychologists, philosophers, and those in the medical profession.

Julie Inness wrote the book on privacy as it relates to intimacy [110]. She disagrees with Fried's understanding of intimacy as a "commodity" which derives its value from its scarcity, and rather proposes that intimate interactions must be *motivated* by liking, love, or care in order to be intimate. As evidence she points to Supreme Court decisions wherein constitutional privacy protection was conferred to issues of the family and sexual health due to the personal, emotional impacts that made those issues intimate. In this way, Inness seems to define privacy as the protection of intimate matters, where intimacy comes from the motivation and

not the behavior itself (e.g., kissing is not automatically intimate). She recognizes that this definition of intimacy is subjective, making legal rulings more difficult.

Privacy might also include solitude, i.e., being physically removed from other people. Solitude is more than a freedom from trespassing; one needn't be at home to desire solitude. Allen [9] includes solitude in her article on privacy and medicine. In the medical setting, the sick often want to be comforted by company, but also to have some time alone. This could be especially true for patients with terminal illnesses, who might want to reflect on their lives and make some important decisions. In such cases we tend to respect their wishes. Allen [9] also mentions "associational privacy," the ability to choose one's own company. She notes that patients do not merely desire intimacy, but rather "selective intimacy" with certain loved ones, and this is an aspect of privacy to consider.

Finally, privacy could be defined in terms of one's personal space or territory. These concepts are found readily in proxemics literature as well as in psychology and ethology (i.e., animal behavior studies) in general. Patricia Newell includes territoriality in her review of *Perspectives on Privacy* [170], although she also cites a study that separates between the two [63]. We have already mentioned that Leino-Kilpi et al. [144] define physical privacy as being over personal space and territory, and Westin also mentions it when he links human privacy ideas with animal behavior in Westin [255].

2.1.1.4 Summary of Definitions.

This concludes our overview of different definitions of privacy from the literature. One begins to see some common themes, but also the scattered nature of the literature. Newell [170] in particular and also Leino-Kilpi et al. [144] do a good job of synthesizing broad sections of what has been written (from psychological and medical perspectives, respectively). See Figure 2.2 for Newell's table of privacy definitions. Newell [170] says, quoting an earlier privacy review, that, "theorists do not agree...on whether privacy is a behavior, attitude, process, goal, phenomenal state, or what." Privacy is mysterious. But we have covered a diverse array of definitions for privacy thus far; DeCew [58] gives a cursory recapitulation:

> Privacy can refer to a sphere separate from government, a domain inappropriate for governmental interference, forbidden views and knowledge, solitude, or restricted access, to list just a few.

2.1.2 Privacy as Described by Prominent Theories

We now present the three main ways we understood privacy for the work presented in this dissertation: *theories* that define privacy (this section), a *taxonomy* that describes its component parts (the next section), and the different ways it is used in everyday conversation by people who are not privacy scholars (Section 2.1.4).

A naïve attempt at a theory of privacy might define it as the extent to which you withhold your personal information from others and keep away from human

TABLE 1
Definitions of privacy found in the literature

Privacy is:	
(a)	not in principle detectable by everyone in the same way (Bailey, 1979)
(b)	the source of activities (Weiss, 1983)
(c)	an instrument for achieving individual goals of self-realization (Westin, 1967)
(d)	a compound of withdrawal, self-reliance, solitude, contemplation and concentration (Chermayeff & Alexander, 1963)
(e)	an attribute of place (Webster, 1979)
(f)	a state of being (Fischer, 1971; Bailey, 1979; Weiss, 1983; Schoeman, 1984)
(g)	a zero relationship between a group and a person (Kelvin, 1973)
(h)	freedom to choose what, when and to whom one communicates (Westin, 1967, Proshansky *et al.*, 1970)
(i)	personal control over personal information (Westin, 1967; Greenawalt, 1971)
(j)	negation of potential power-relationships (Kelvin, 1973)
(k)	the right to be left alone (Cooley, 1880; Brandeis & Warren, 1890)
(l)	control of personal space (Hall, 1969; Canter & Canter, 1971; Canter, 1975; Gold, 1980; Fisher *et al.*, 1984; Duvall-Early & Benedict, 1992)
(m)	a central regulatory process (Altman, 1975)
(n)	a voluntary and temporary condition of separation from the public domain (Newell, 1992)
(o)	a valued commodity (Loo & Ong, 1984)
(p)	a state in which persons may find themselves (Velecky, 1978)
(q)	a value that should be considered in reaching legal decisions (Gavison, 1984)

Figure 2.2: Patricia Newell's summary table of privacy definitions from the (largely psychological) literature [170].

contact. Privacy in this sense would be most effectively achieved by the hermit living in a remote hut, shunning all visitors. The problem with this view is that we often *want* to confide in other people and be part of human society. In fact, humans want to share their lives with others—just not too much.

Adam Moore offers an alternative definition to account for this objection: according to him, privacy is "control over access to oneself and information about oneself" [164]. This is a "control-based" definition of privacy, in which it doesn't matter whether somebody accesses you or your information, but rather whether you can control that access. Control-based definitions account for situations in which someone invites others into his or her close company, or willingly gives out personal information. These actions would violate privacy if privacy is the state of being let alone, or of having all your personal information kept to yourself. But authors holding to control-based definitions of privacy maintain that the person in question is still in control, so there's no violation; this especially makes sense in the legal context.

Irwin Altman offers a more sophisticated theory. He defines privacy as a *boundary regulation process* wherein people try to achieve their ideal privacy state by using certain *mechanisms* to regulate interaction with others [10]. Notice how this definition allows privacy to sometimes mean *more* interaction with others, and sometimes *less* interaction; successfully switching between the two is the key. Along these lines, Altman calls privacy a *dialectic process*, i.e., a contest between two opposing forces—withdrawal and engagement—which alternate in dominance. Hence, privacy to Altman is *dynamic* in that the desired level of engagement

changes over time for a given individual. This theory is necessary for understanding Altman's discussion of personal space, territoriality, and crowding.

Finally, Helen Nissenbaum's approach to privacy, which she calls "contextual integrity," focuses on the idea that different norms of information gathering and dissemination are observed in different contexts [172]. Privacy is violated in a given context when the norms for information gathering or dissemination within that context are broken. Nissenbaum argues that some scenarios, especially public surveillance, are intuitively felt by many to be potential privacy violations, and that while U.S. legal policy overlooked these scenarios (at time of writing), "contextual integrity" does a better job of accounting for our intuitive concerns [172].

2.1.3 Privacy as a Taxonomy of related Constructs

Whereas theories about privacy try to define it as a whole, a taxonomy describes its parts and shows how they are related to each other. There are different types of privacy—e.g., regarding personal information vs. social boundaries. This section presents a breakdown of privacy into its different sub-concepts that has been organized into a hierarchical taxonomy. This taxonomy attempts to organize all the different types of privacy mentioned by privacy scholars into a single structure. This is the second of three ways we understood privacy in our work.

A paper describing this taxonomy has been placed on arXiv [196]. This section draws text from that paper.

2.1.3.1 A Taxonomy of Privacy Constructs for Human-Robot Interactions

This section lays out our taxonomy of privacy constructs and summarizes the key literature behind it. Definitions of terms are to be found via the references where not defined hereafter. The taxonomy is as follows:

Privacy (see Leino-Kilpi et al. [144] for subdivision)

1. Informational (see Solove [223] for subdivision)
    (a) Invasion
    (b) Collection
    (c) Processing
    (d) Dissemination
2. Physical
    (a) Personal Space [255]
    (b) Territoriality [255, 170, 39] (see Altman [10] for subdivision)
        i. Intrusion
        ii. Obtrusion
        iii. Contamination
    (c) Modesty [9]

3. Psychological

    (a) Interrogation [255]

    (b) Psychological Distance [95]

4. Social

    (a) Association [9]

    (b) Crowding/Isolation [10]

    (c) Public Gaze [17]

    (d) Solitude [9] (see Westin [255] for subdivision)

    (e) Intimacy

    (f) Anonymity

    (g) Reserve

2.1.3.2 The Literature behind the Taxonomy

We recommend the Stanford Encyclopedia of Philosophy article on privacy by Judith DeCew as a comprehensive guide to the definition of privacy [58], especially in law and philosophy. Most of the references in this section we owe to the bibliography from that article.

1-4 Leino-Kilpi et al. [144] divide privacy as follows:

1. Informational privacy, over personal information

2. Physical privacy, over personal space or territory

3. Psychological privacy, over thoughts and values

4. Social privacy, over interactions with others and influence from them

1.a-d Informational privacy refers to privacy concerns about personal information. In 1960, William Prosser divided (informational) privacy into four parts. His formulation continues to be referenced today. Briefly, Prosser divides (informational) privacy into intrusion, public disclosure, false light, and appropriation. These mean the following. First, intrusion into one's private affairs includes trespassing, search, and remote intrusion such as wire tapping. Second is public disclosure of private facts. Third is publicly portraying the victim in a false light, e.g., by misattributing to the victim a statement or opinion. Fourth is appropriation, or pretending to be the victim for one's own advantage. Daniel Solove has constructed a taxonomy of privacy concepts based on Prosser's formulation. It is shown in Figure 2.1 as a general overview of informational privacy concerns. We use the highest level of Solove's hierarchy for 1.a-d.

2.a-b Privacy could be defined in terms of one's personal space or territory. These concepts are found readily in proxemics literature as well as in psychology and ethology (i.e., animal behavior studies) in general, but are not often connected with privacy. Patricia Newell includes territoriality in her review of *Perspectives on Privacy* [170], although she also cites a study that separates between the two [63]. Leino-Kilpi et al. [144] define physical privacy as being over personal space and territory, and Westin also mentions it when he links human privacy ideas with

animal behavior [255].

Social psychologist Irwin Altman pulls together the related concepts of privacy, personal space, territoriality, and crowding [10]. His book, along with Burgoon's article [39] (discussed below), is a good foundation for *environmental and spatial* factors related to privacy.

Judee Burgoon presents a communication perspective on privacy, including territoriality, in a broad survey [39]. She argues that more "physical" privacy could consist of blocking more communication channels, including sight, sound, and even smell (e.g., the smell of food being cooked next door). We would add further channels enabled by technology: phone calls, text messages, Facebook posts, and the like. Alternatively, Burgoon writes that to have more territory, higher-quality territory (e.g., better-insulated walls), and more unquestioned control over that territory is to enjoy more physical privacy.

2.c Allen lists *modesty* as an important physical privacy concern in medical settings, especially from the philosophical standpoints of Christian ethics and virtue ethics [9]. Modesty may drive patients to request same-sex or even same-race doctors.

3.a According to Westin's account of privacy in U.S. law, the right to privacy swelled in the late 1900's [255]. The Supreme Court continued to try cases in which new technologies created privacy concerns beyond physical entry and tangible items. According to Westin, new protections included "associational privacy" over group memberships (this is distinct from 4.a) and "political privacy" over unfair questioning on account of political positions.

3.b *Proxemics* can include psychological distance as well as physical distance (see Hall [95] cited by Mumm and Mutlu [165]).

4.a and 4.d Privacy might also include solitude, i.e., being physically removed from other people. Solitude is more than a freedom from trespassing; one needn't be at home to desire solitude. Anita Allen includes solitude in her article on privacy and medicine [9]. In the medical setting, the sick often want to be comforted by company, but also to have some time alone. This could be especially true for patients with terminal illnesses, who might want to reflect on their lives and make some important decisions. In such cases we tend to respect their wishes. Allen also mentions "associational privacy," the ability to choose one's own company [9]. She notes that patients do not merely desire intimacy, but rather "selective intimacy" with certain loved ones, and this is an aspect of privacy to consider.

4.b Altman calls both crowding and isolation failures to regulate the amount of interaction with others [10]. It may seem odd to call social isolation a privacy issue, but it is a logical conclusion from within Altman's theory of privacy (see Appendix).

4.c Lisa Austin offers a more nuanced definition of privacy: freedom from "public gaze" [17]. She argues that this updated definition deals with the problem of new technologies to which older definitions of privacy do not object. In particular, Austin is concerned about cases wherein people know they are under surveillance, about the collection of non-intimate but personal information (e.g., in data mining), and about the collection of personal information in public. She claims that other, older definitions of privacy do not agree with our intuition that these tech-

nologies (could) invade our privacy by denying us our freedom from "public gaze."

4.d-g Alan Westin lists four different states of privacy: solitude, anonymity, intimacy (i.e., being alone *with* someone), and reserve (i.e., keeping to oneself) [255].

2.1.3.3 Measuring Constructs from the Taxonomy

This taxonomy takes the broad concept of privacy and breaks it into more specific constructs. We have split the single trunk into what we see as its main branches, and some of those branches have also been shown to fork off, too. To study privacy in human-robot interaction (e.g., in human-subject experiments), we need the leaves of this privacy tree. Unlike the trunk and branches, the leaves are no longer abstract constructs; instead, they are concrete measures. For example, one operationalization of personal information collection (1.b) would be whether someone knows your social security number—a simple, binary measure. Other measures might be contextual, e.g., given that you are alone in a room with a PR2 robot staring at you, do you feel comfortable changing your shirt? This comfort level, a proxy for modesty (2.c), could be measured, for example, by a questionnaire. All such measures would tap the extent to which a person's privacy has been preserved or violated.

2.1.4 Privacy as a Word with Multiple Folk Definitions

So far we have presented scholarly definitions of privacy. Our third way of understanding privacy is by considering how it is used by non-scholars in everyday conversation. I.e., we must remember that "privacy" is a word in the English language. A word can have multiple meanings or be used differently by different individuals or cultures. This perspective on "privacy" is important when we brief our study participants, write questionnaires, and analyze what participants say. Asking them, "how concerned are you about your privacy when this robot is in the room?" could be interpreted in multiple ways. For example, in everyday conversation the same person could say, "I went to my room for some privacy" (i.e., solitude) and, "I adjusted the privacy preferences on my Facebook account" (i.e., information privacy)—the same word is used to mean two different things.

We do not believe, however, that it is always necessary to avoid using the word "privacy" because of its ambiguity when talking to participants. Rather, we must be clear about which sense of the word we mean, e.g., "privacy, by which I mean control over who can start a conversation with you," perhaps even operationalizing a construct into something measurable, e.g., "access privacy, which here means the fraction of daily conversations that you did not want to have." The "fraction" in the latter example is relatively unambiguous compared to the word "privacy" or even "access privacy;" researchers should favor such language to avoid misunderstandings. When analyzing spoken or written responses, researchers should be skeptical of what a participant means by "privacy" and other words with multiple

meanings. Searching the context for clues or—even better—asking the participant for clarification can help in these situations.

2.2 Privacy Scholarship: a Survey across different Disciplines

We now present a review of privacy scholarship across a variety of disciplines. The *history* of privacy in philosophy and U.S. law is given first as a sort of backdrop. Psychology, anthropology, and several other social sciences are discussed next along with medicine as the main sources of *scientific research* on privacy. Finally, technology is discussed as an *application domain* for privacy.

2.2.1 History: Privacy in Philosophy

Privacy has a long but scattered history in the field of philosophy. Several philosophers, such as Solove, Thomson, and DeCew, have already been mentioned above as they contributed to defining privacy. This section includes some older thinkers, and presents some useful distinctions that go beyond just defining privacy. Here again, Judith DeCew's article in the Stanford Encyclopedia of Philosophy is our guide [58].

Aristotle wrote the earliest extant philosophy of privacy. Remember, however, that modern usage of the word "privacy" is different than ancient usage. The Greek concept of privacy was a distinction between the affairs of the *oikos* (household) as separate from the *polis* (city-state) [8]. The former situation had free males ruling

collectively over the city-state, whereas in the latter they ruled "despotically" over the household. So when, in Book II of his *Politics*, Aristotle asks whether property ought to be held "privately" or "in common," he is asking whether each individual man should own it or whether all the men should own it together [13]. To this question he answers, "privately." Aristotle argues that, if men own things privately, they will take better care of them and enjoy them more. Furthermore, sharing will take place naturally (an optimistic claim!), and all the evils commonly blamed on the paradigm of private property are actually due to human wickedness. Hence, Aristotle is an early proponent of our Western ideal of private ownership, which undergirds certain aspects of privacy.

John Locke was an Enlightenment philosopher who greatly influenced American political thought. In his liberal philosophy, Locke was a promoter of human rights. Adam Moore has written a book on the Lockean idea of intangible property, which includes things like intellectual property which we do not purchase but nonetheless possess by right [163]. According to Locke, things that we create are ours as the right of the laborer, "at least where there is enough and as good left for others" [163]. In other words, our ideas or self-expressions can be ours by property rights so long as they aren't significantly depriving anyone else of similar rights. Applying this rule can be tricky in the legal context.

Modern philosophers joined the conversation about privacy that started around the beginning of the 20th century. They especially help to subdivide the topic into clear categories. For example, they distinguish between *descriptive* and *normative* accounts of privacy, i.e., between describing what privacy covers and examining

why it's important as a value or right. The rest of this section is divided according to that distinction; we start by looking at some *descriptions* of privacy, then move to some work on the *normative* side.

A key contribution by philosophers to descriptive privacy is the coherentist-reductionist debate [58]. Schoeman [214] coined the words "coherentist" and "distinctivist" to represent separate groups of people, but we will take them together as a single group that talks about privacy as an independently important idea that cannot be *reduced* to other ideas—"reductionists" hold that other view [58]. We think that taking the reductionist view here would contradict popular opinion and shatter this one paper into many, so we shall remain coherentists. Nevertheless, it is valuable to understand the reductionist viewpoint, stated most famously by Judith Thomson.

In her 1975 article, *The Right to Privacy*, Thomson does not debunk privacy altogether but rather calls it a "derivative" right, i.e., a right based on other rights [241]. The argument runs as follows. First, we have rights over not just our property but (even more so) over our own bodies. What might these rights of the body include? Here Thomson cites what she believes to be the earliest statement of body rights: "A Definition of Privacy," written in 1974 by Professor Richard Parker. Parker says the following:

> Privacy is control over when and by whom the various parts of us can be sensed by others. By 'sensed,' is meant simply seen, heard, touched, smelled, or tasted. By 'parts of us,' is meant the parts of our bodies, our voices, and the products of our bodies. 'Parts of us' also

> includes objects very closely associated with us [e.g., possessions].

Thomson's examples of body rights are the right to not be looked at and the right not to be listened to. She is quick to state that these rights are not always *claimed*; rather, we usually *waive* them implicitly when we go out in public, perhaps whistling as we walk. But they become very apparent in some cases: Muslim women would surely claim the right not to have even a bare knee seen by a stranger, for example.

Surprisingly, Thomson does not throw privacy out completely here. Instead, she defines each privacy right as some other right with a personal component. For example, someone reading my book violates my right to it as my property, whereas someone reading my diary violates the *same* right, but with regard to my personal information, from which arises an additional right to privacy. Or consider the case of torture: torturing me to extract some random fact, like the capital of the state of Oregon, only violates my body right to go unharmed by others. Torturing me for *personal* information, however, violates the same body right *and* my further right to privacy. The key here is the *direction* of the logic: my right not to be harmed for my personal information does not come from my right to privacy, but rather the other way around.

Most authors, however, talk about privacy as distinct and not (wholly) derived from other rights. On that note, we now consider some *normative* accounts of privacy; that is, some different conceptions of why privacy is important. Why is it good or useful to value privacy? For Bloustein [32], privacy is about the safety and freedom of each person's individuality and dignity. Fried notes its necessity for

friendship and trust [78]. Inness among others says it is necessary for intimacy with others [110]. James Rachels, besides making several other cogent contributions that have been kept from this paper for brevity, addresses the question of why we need privacy if we aren't doing anything embarrassing [187]. He answers that the presence of other people influences our actions, so being unable to control one's company is actually a loss of autonomy. Hence, privacy is a concern even when we're behaving ourselves—in fact, inasmuch as we feel pressured to "behave," privacy includes the freedom to escape such a situation.

2.2.2 History: Privacy in U.S. Law

The story of privacy law in the United States is in many ways the crux of the history of privacy. What began as essentially a set of property rights has grown at both the state and federal levels to include rights to confidentiality and about intimate personal decisions. This section surveys the major contributions in chronological order, beginning with the ratification of the Bill of Rights in 1791.

The Fourth Amendment to the U.S. Constitution begins as follows:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated... [1]

...to which the reader should silently add, "by agents of the *government*." The Bill of Rights was drafted in order to protect the rights of U.S. citizens from an unlimited federal government; it was only by the doctrine of "incorporation"

that, starting in the 1920s, courts began holding state governments accountable for these rights as well. History lessons aside, the Fourth Amendment has been one foundation, although not the only one, of court rulings about privacy rights in the United States.

In 1967, Alan Westin published *Privacy and Freedom*, the result of over four years of effort and study by a group of researchers [255]. His book includes a chronology of privacy law in the United States which we will follow here with some interjections from other authors. Westin begins with America as a new nation, framed by the Constitution. Compared to Europe, America had much in the way of privacy law before the Civil War [255]. American politics were founded on the worth of the individual, limited government, property rights, and the liberty to do what one will with his property. The First Amendment conferred the freedom to speak, but also to keep silent. Anonymous publication of one's opinions was allowed, and police surveillance of public meeting places, as was prevalent in Europe, was expressly forbidden. Justice Story held that the Third Amendment (against housing troops in private homes) was meant to secure, "that great right of the [English] common law, that a man's house shall be his own castle, privileged against all civil and military intrusion" [255]. There was also judicial precedent (part of common law) that forbade nuisances, trespassing, and eavesdropping. Trademarks, corporate and government secrets, and letters all enjoyed general protection.

It was technology that began causing problems from the legal perspective [255]. The telephone was invented in the 1880s, followed closely by wire-tapping within

a decade. Similarly, the microphone was invented in the 1870s, followed by the dictograph recorder in the 1890s. For capturing images, "instantaneous photography" became possible with the Kodak camera in the 1880s, and by 1902 *The New York Times* reported on what we would call the first paparazzi: "kodakers" waiting to capture photos of important people [255]. It is not a coincidence that this is when the Supreme Court and legal scholars began to treat seriously the expansion of American privacy rights.

We have already mentioned the true beginning of the privacy conversation in America: the short article written in 1890 by Samuel Warren and Louis Brandeis entitled, "The Right to Privacy" [251]. The occasion was a humorous one. Warren's wife was among the social elite and her parties tended to be written up in the newspaper; eventually, Mr. Warren got fed up and joined with Brandeis to write the article [185]. Warren and Brandeis cite Judge Cooley's right "to be let alone" in the context of increasing invasions by journalists of Americans' private lives [251]. They argued that both the intimate and the mundane details of one's life are protected, along with one's thoughts, sentiments, and emotions, until "published" to others. Violations are to be redressed by tort (i.e., monetary compensation) or by a court injunction (i.e., restraining order).

Prosser [185] reports that the article by Warren and Brandeis was argued over for about 30 years. By the 1930s, its arguments began to be accepted by the courts, and most states had privacy regulations along the lines of Warren and Brandeis by 1960. Prosser reviewed the over 300 privacy cases on the books by 1960 and reports on the emergence of four separate torts over the years. We have already discussed

them in Section 2.1.1.1: invasion, public disclosure of private facts, false light in the public eye, and appropriation of someone's name or likeness. We now continue with Westin's account [255]. The right to privacy swelled in the late 1900s. The Supreme Court continued to try cases in which new technologies created privacy concerns beyond physical entry and tangible items. Other new protections included "associational privacy" over group memberships and "political privacy" over unfair questioning on account of political positions. Anonymity in public opinion was upheld in *Talley v. California, 1960* [255]. Privacy of the body was upheld in a 1964 case wherein the "reasonableness" of certain compulsory medical examinations were in question [255]. Most importantly, and as discussed above in Section 2.1.1.2, "constitutional privacy" emerged in 1965 with the *Griswold v. Connecticut* decision [255].

In 1948, the Universal Declaration of Human Rights was adopted by the United Nations General Assembly [239]. Although it came before the rise of "constitutional privacy" in America, this document communicated an updated understanding of privacy rights to the world:

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks. (Article 12)

The words "interference" and "privacy" seem to be quite general here; we see this statement as a pretty broad dispensation of privacy rights.

A number of authors have written about privacy more recently (say, since 1980). We have already discussed several in Section 2.1.1 [180, 17, 223] that contributed to *defining* privacy in the legal domain. Beginning in the 1970s, however, feminist criticism of traditional understandings of privacy arose [8]. Anita Allen, for one, reminds privacy scholars that home life is not automatically good; privacy can hide abuse, subordination, and isolation, especially for women [8]. Allen believes that we can salvage privacy as a society, however, which is good news seeing as constitutional privacy rights about abortion and contraceptive use are especially important for protecting women. Her argument seems a bit conflicted in places, but underlines that women can be especially affected by the way we implement privacy as a society.

2.2.3 Research: Privacy in the Social Sciences

Various social sciences—anthropology, sociology, psychology, economics—have studied privacy in humans. That being said, one could also argue that the roots of human privacy values can be seen in animal behavior—that is, in ethology. Alan Westin states that almost all animals seek privacy, either as individuals or as small groups [255]. It is from animals that we get the idea of *territoriality*, or the defense of one area against intrusion by others of one's own species (which raises questions, by the way, about what species a robot might be—see Section 7.1.1.7). Westin reports three types of spacing observed between animals: personal distance between individuals (e.g., between birds on a wire), social distance between groups, and

flight distance at which an intruder causes fleeing [255]. At the same time, animals often gather in large crowds, and it seems that at least part of the reason is the mere social stimulation. Animals, like humans, seem to live in a tension between privacy and sociality.

The English words "private" and "privacy" come from the Latin adjective *privatus*, meaning "set apart, belonging to oneself (not to the state), peculiar, or personal." This is used in contrast to *publicus* and *communis*, which have the expected meanings. *Privatus* is itself the past participle of *privare*, a verb meaning "to separate or deprive," which in turn comes from the adjective *privus*, meaning "one's own, individual" [4]. Several modern languages lack a word with the exact sense of English "privacy;" these include Arabic, Dutch, Japanese, and Russian [170]. French has some related words, but none covers the concept in general. Japanese has four related concepts with similar meanings to "private life," "freedom," "solitude," and "secrecy" in English [170]. Therefore, having a single word that is even roughly equivalent to the English "privacy" is *not* universal across languages.

Is privacy-regulating behavior peculiar to Western cultures, or even to American culture? Irwin Altman investigates in his study entitled, "Privacy Regulation: Culturally Universal or Culturally Specific?" [11]. He hypothesizes that privacy is a universal human need, but that different cultures have different privacy regulation mechanisms. Altman surveys two types of cultures: those with apparently minimal privacy and those with apparently maximal privacy. If mechanisms exist to regulate privacy in those extreme cultures, he argues, then privacy must be

universal despite all appearances. This was in fact the case. For example, the Mehinacu culture in Brazil appears to have almost no privacy; housing is communal, people enter huts without announcement, and each person's actions and whereabouts are generally known [192]. Upon a closer look, however, we see mechanisms for regulating privacy. It is permissible to leave the village for days at a time, or to walk various secret paths alone and unaccounted for. Lying was used regularly to avoid revealing too much while living in such close quarters. Some privacy-related behaviors were humorous, especially that of the Sapps people of Northern Europe. These are reindeer herders who live in tents, and while anyone could make oneself welcome in anybody's tent, a disgruntled occupant would often feign sleep to signal for privacy [179].

We move now to studies of human privacy in the modern, developed world. Privacy during childhood is a good place to start. Zeegers et al. [261] found that 58 of 100 three-, four-, and five-year-olds said they had a special place at the daycare center that belonged only to them. When confronted with reduced time and space for privacy, children in daycare can become territorial, more defensive of their personal space, and more attentive of their physical privacy [113]. Newell [170] mentions the importance of "refuge" to autistic children especially; she even reports that some children use poor behavior to get put in isolation.

What about privacy norms in adults? Patricia Newell's study of American, Irish, and Senegalese subjects (mostly students) found that they usually sought privacy when sad or tired, or to concentrate, and felt relaxed and refreshed afterwards [169]. Friedman et al. [79] studied the difference in privacy judgments

between the so-called "watcher" and the person being "watched" in a public setting. The results revealed a gender difference: women were significantly more likely to be surprised or troubled by the idea of being watched via a video camera. Also, men were less privacy-sensitive as watchers than they were when watched.

Acquisti and Grossklags [5] study the dichotomy between privacy *attitudes* and privacy *behaviors*. Sometimes, even people who value their privacy don't seem to act to protect it. The authors hypothesize three reasons for this dichotomy. First, people have incomplete information when they make privacy decisions. Second, humans have only "bounded rationality," so we can't make perfectly rational decisions if there is too much information to consider. Third, we have certain systematic biases. For example, we lack self-control and opt for instant gratification even when it's sub-optimal in the long run. The results of a 119-subject survey supported all three hypothesized factors. For example, 44% of subjects showed a time-inconsistency bias by discounting future payoffs. This suggests that people cannot be expected to make fully rational decisions about privacy, even if they value it dearly.

Another study by Acquisti and his colleagues assessed the *subjective value* of privacy [6]. They draw an important distinction between the willingness to accept (WTA) payment to give up one's privacy and the willingness to pay (WTP) for privacy protection. Economists have found that people are willing to accept only a very high payment to give things up, perhaps due to "loss aversion" [6]. The results gathered by Acquisti et al. [6] showed five times higher WTA than WTP in a privacy-related context; subjects were unwilling to make even small monetary

sacrifices to gain privacy protection. They also found that valuation of privacy depends on how you ask. It depends very much how much privacy protection a person is given to start with, and whether you propose to take some of it away or to provide more.

2.2.3.1 Personal Space, Territoriality, and Environment.

Social psychologist Altman [10] pulls together the related concepts of privacy, personal space, territoriality, and crowding. His book, along with work by Burgoon [39] and Sommer [225], discussed below, is a good foundation for *environmental and spatial* factors related to privacy. Altman's theory defines privacy as a *boundary regulation process* wherein people try to achieve their ideal privacy state by using certain *mechanisms* to regulate interaction with others. Notice how this definition allows privacy to sometimes mean *more* interaction with others, and sometimes *less* interaction; successfully switching between the two is the key. Along these lines, Altman calls privacy a *dialectic process*, i.e., a contest between two opposing forces—withdrawal and engagement—which alternate in dominance. Hence, privacy to Altman is *dynamic* in that the desired level of engagement changes over time for a given individual. This theory is necessary for understanding Altman's discussion of personal space, territoriality, and crowding.

To Altman, personal space and territoriality are two *mechanisms* for achieving a desired privacy state. Another example is speech. Reviewing the literature at the time, Altman reports that there are individual differences in the use of personal

space, but it's hard to confidently state clear differences between men and women or between different cultures. It does seem that people maintain less personal space with people they know and like, as well as in informal settings. Personal space is also relationship-dependent; being too close is an invasion for strangers and some interactants, but excessive *distance* can be undesirable between intimates. As for territoriality, Altman notes that the phenomenon in animals is related in some ways, but also markedly different in other ways. He lists three kinds of territories: primary (e.g., a bedroom or home), secondary (e.g., a local bar), and public (e.g., a bus seat). The three kinds vary in the extent to which *markers* and other strategies can be used to claim and defend space. Altman also distinguishes from outright *intrusions* both *obtrusions*—defined as "excessive use" of the space, such as being too noisy—and *contaminations* of the space. Altman recommends longitudinal studies in natural environments for future research on territoriality.

Altman distinguishes *crowding* from *density*. Crowding is when too much social interaction occurs, whereas density is a quantitative measure of persons per area, per room, per home, etc. The opposite extreme from crowding is social isolation; both are bad and indicate failure of the regulatory mechanisms that enforce a person's privacy. Altman presents results of laboratory experiments on the effects of crowding on people, as well as various mechanisms for coping with crowding.

Judee Burgoon presents a communication perspective on privacy, including territoriality, in a broad survey [39]. She argues that more "physical" privacy could consist of blocking more communication channels, including sight, sound, and even smell (e.g., the smell of food being cooked next door). We would add further

channels enabled by technology: phone calls, text messages, Facebook posts, and the like. Alternatively, Burgoon writes that to have more territory, higher-quality territory (e.g., better-insulated walls), and more unquestioned control over that territory is to enjoy more physical privacy. She goes on to list many aspects of physical privacy: in spatial intrusions, the probability of violation, humanness of the violator, and relationship with the violator; in social interactions, control over who the interactants are, the frequency and length of the interaction, and the content of the interaction; also, formality allows interactions to be kept short and impersonal, whereas a more private state is characterized by greater rule-breaking, "backstage" behavior, and freedom to engage in emotional release.

Burgoon cites Robert Sommer's book on personal space with respect to the possibility that an intruder would be treated as a "nonperson" [225]. Sommer observes that inanimate objects like trees and chairs, as well as pets in some circumstances, are not treated as intruders at all. This same phenomenon can occur with people. For example, a cubicle dweller might continue a sensitive phone conversation even as the janitor enters and begins emptying the trash. In certain social contexts, such as subway trains and sporting events, it can be socially normal to treat others as nonpersons.

Sommer also studied territorial markers. To effectively protect an area, markers need to either be an explicit sign (e.g., a "Reserved" sign) or to have intrinsic value (e.g., coat, purse, umbrella); litter doesn't count. Valid markers can be very effective at protecting a spot, especially during low room density. As more people enter the space, markers are more likely to be ignored, and others in the room

might be asked to confirm whose territory that is.

We now return to Burgoon's list of privacy *mechanisms* in Altman's sense (see discussion of Altman [10] above), clearly influenced by her expertise in communication. She lists six categories of mechanisms. First is environment and artifacts, which is further broken into architectural design, artifacts and furnishings, and gatekeepers (e.g., receptionists effectively guard the rest of the building). Factors like the size of each area, the blockage of senses between areas, and other cues that say "socialize here" or "don't" all impact the privacy of a space. Second is personal space and touch, which includes interpersonal distance, seating positions at tables, body orientation, degree of sideways or backwards lean, and use of physical touch. Third is chronemics, or the usage of time. For example, people could use the same space but at different times, avoid social areas during peak usage hours, or even declare different functions for a shared space at different times. Fourth is kinesics and vocalics, or cues from the body and voice. Examples of "exclusion cues" include "body blocks" or closed postures in a dyad, saying "go away," silence, or avoiding eye contact so others can't begin visual communication. "Affiliation cues" include smiling, relaxed postures, greater mirroring, and warm vocal tones. Fifth is physical appearance and attire, and sixth is verbal mechanisms. Numerous verbal mechanisms exist. The linguistic features of speech, such as tense, use of personal pronouns, use of negation, and ambiguity can regulate privacy. So can the degree of self-disclosure and formality, changing the topic, brevity or verbosity, and direct references to one's possessions, territory, or rights. Finally, there is the idea of "linguistic collusion"—using in-group language to exclude others [39].

We will mention several field studies in this area. First, Becker and Mayo [25] note the possibility of confusing the concepts of personal space and territoriality. They study whether what Sommer and Becker [226] call territorial behavior in a cafeteria setting is actually closer to Hall's concept of personal distance [95]. Is using markers to reserve a spot at a table really just a mechanism for maintaining a comfortable personal distance? If so, the person who marked the space will be just as likely to move to a new spot upon intrusion than defend that spot in particular. A study of 48 unsuspecting university students supported this hypothesis. The authors argue that the active construct in that scenario was personal distance, not territoriality.

Walden et al. [250] study how incoming freshmen at an American university cope with different levels of crowding. The authors measure both objective and subjective crowding (what Altman calls density and crowding, respectively; see above). They also highlight the difference between one's *expectation* of achieving the desired privacy state and the *value* or pay-off of achieving that goal. For the study, students were assigned either a two-person or a three-person dormitory room and surveyed about their experiences. The results showed an apparent difference between the way males and females cope with crowding, but were limited by a low sample size, especially of male subjects.

Sebba and Churchman [216] interview 45 Israeli families with two to four children living in apartments of almost identical size and layout. Based on their results, they classified territorial areas into four categories. The first three were called individual, shared, and public areas. Single-occupancy bedrooms were indi-

vidual territorial areas, multiple-occupancy bedrooms were shared, and most living rooms were public. The final category they called jurisdiction areas, wherein one person has jurisdiction but the space is used by everybody. About half of the kitchens were defined as jurisdiction areas—they were used by everybody, but belonged to the mother. Additional research is needed to determine whether these findings generalize to other demographic groups and types of spaces.

2.2.3.2 Online Privacy.

Privacy on the Internet is a bit of a different animal. The user can browse as an anonymous, disembodied agent, and is confronted with compelling advertisements and endless information. In a study of online interaction in general, subjects were guided by an online avatar named "Luci," who asked them personal questions to help them pick out some things to buy [30]. Some questions were purely relevant to the purchasing task, while others were more prying. In general, subjects gave away a lot of information, and again this was true even of those who self-reported as valuing their personal information.

Paine et al. [178] studied *why* people (don't) act to protect their privacy online. Five hundred and thirty people from various countries responded to an open-ended survey administered by an automated instant-messaging bot. When asked why they were (not) concerned about privacy online, those concerned cited viruses, hackers, etc., whereas those not concerned cited their own competency, protective software, apathy, and a lack of valuable information to hide. Those who reported

not taking privacy-protective actions cited as reasons apathy, a lack of felt need for protection, and (this was troubling) not knowing *how* to get effective privacy protection online. Again, the questions were open-ended and subject to respondents' interpretations of "privacy online," but the results enumerate some issues that need addressing in online privacy.

Of special interest to this paper is the development of privacy metrics and evaluation methods, since those could possibly be adapted for use in human-robot interactions (see Section 7.1). Buchanan et al. [37] construct a new online privacy measure (namely, a questionnaire) and test it against two popular privacy measures from the previous literature. They propose three privacy metrics to be measured by a 28-item questionnaire: General Caution (when using the Internet), Technical Protection, and Privacy Concern. These metrics were validated both as constructs and against two other scales: Westin's Privacy Segmentation [98] and the IUIPC [156]. Over 1000 respondents took all three measures, and results were significantly correlated in general.

2.2.4 Research: Privacy in Medicine

The field of medicine is aware of the importance of privacy. This is evidenced by the *Declaration on the Promotion of Patients' Rights in Europe*, which includes respect for privacy as 1 of its 6 human rights [239]. The *Declaration* frowns on any medical treatment that cannot be performed with respect for the patient's privacy.

Allen [7] lists three main uses of the term "privacy" in the healthcare domain: physical privacy, informational privacy, and decisional privacy. The third usage—decisional privacy—is especially salient in healthcare. Allen defines decisional "privacy" as the freedom to make one's own decisions (here, about medical treatment) and act on them without undue outside influence. Careful readers will notice that decisional "privacy" is the same as constitutional privacy, discussed above in Section 2.1.1.2. Allen places quotation marks around the word "privacy" in decisional "privacy" because of some controversy over whether decisional "privacy" should be called privacy at all, but rather liberty, freedom, or autonomy [7]. In a different article, Allen [9] lists *modesty* as an important physical privacy concern in medical settings, especially from the philosophical standpoints of Christian ethics and virtue ethics. Modesty may drive patients to request same-sex or even same-race doctors.

Things we consider common etiquette, such as knocking on a door before entering, can be supportive of physical privacy [144]. Also, technology is causing problems in the medical as well as the legal realm; a survey in a Finnish hospital revealed that, "Only 30% of 166 respondents, however, believed that their data was safe in the hospital's computer system" [144]. In a 2005 study of an Australian emergency department, 105 of 235 survey respondents reported a privacy breach, defined as either personal information being overheard or private body parts being seen [121]. Influencing factors included the length of stay and whether patients were separated by solid walls or just curtains.

Applegate and Morse [12] have published a study of privacy in a nursing home

for Canadian war veterans. The authors focused on the relationships between the residents, between the staff members, and between residents and staff. Relationships were categorized by how relational others were treated: as friends, strangers, or inanimate objects. This last phenomenon is understood as *dehumanization*, and took several different forms. Privacy was violated most often for those who were dehumanized. Dehumanization was more common towards the less mentally competent residents; other residents would sometimes push them out of the way, and staff members might administer their medicine forcibly and without verbal acknowledgement. This study highlights how privacy-relevant behavior is subject to other factors—here, the mental act of dehumanization—that might not be expected by the privacy researcher but nevertheless drastically affect the situation. Moreover, here we see that there are places wherein privacy is not a fringe concern, but rather an everyday concern, a central quality of life issue like it is in a prison. These may be edge cases, but they are nonetheless motivating.

2.2.5 Applications: Privacy in Technology

Video media spaces (VMS) connect people separated by distance with video channels. Mediated interactions differ from face-to-face interactions in several key ways; Boyle et al. [36] present a vocabulary for understanding these differences. *Disembodiment* is the stripping of context (e.g., at home, hard at work) from the interaction. *Role conflict* is when a media space places someone in several disparate contexts at once, as in the familiar case of working from home. *Dissociation* is

the decoupling of one's body and identity from one's actions, as with a remote operator of a robot. Social signals that are lightweight for in-person interactions are more difficult in a media space. For example, signalling availability (regulating *solitude*) is done with nuanced facial expressions, tones of voice, hand and posture signals, and environmental cues like a door leaned shut. This is all possible in video media spaces, but is currently awkward instead of natural and blunt instead of nuanced.

Boyle et al. [36] continue by arguing that privacy risks are unavoidable with technology. A privacy *risk* has both *probability* and *severity*. Privacy risks are only worth it if counterbalanced by some reward; the technology must provide a comparable benefit. Just as for other personal rights, privacy rights need protection by *policing*, which includes real *punishments* for violations and warrants the formulation of privacy *rules*. This is especially true for situations with an imbalance of power, such as when one person can access (i.e., see and hear) another person's space via a one-way connection. We can restore balance through *reciprocity*, which is when person A can do to person B what B can do to A.

The authors also distinguish between *access control* over who can use the media space and *content control* over what users can see and hear [36]. Content control is often provided by filters. For example, the eigen-space filter by Crowley et al. [53] ensures that only images from a socially-acceptable set will be displayed. Extreme examples of filtering include the availability indicators (e.g., green means available, red means busy) used in instant messaging applications. This brings up concerns about the minimum amount of *fidelity* needed for some task, and also

about *data integrity*, i.e., that the video feed is faithfully modified before it reaches the recipient. Finally, the authors mention evidence that mandatory media space usage causes social changes among the users over time [36]. One could speculate that robot usage will cause even more drastic changes.

Privacy is also a topic of conversation within the subdiscipline of *ubiquitous computing*. Weiser [254] defines ubiquitous computing, or "ubicomp," as, "the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user." Hong and Landay [104] claim that "privacy is easily [ubiquitous computing's] most often-cited criticism." One major principle that has emerged from this conversation is to consider privacy issues during product design instead of just making rules to fit the products. Bellotti and Sellen [27] introduce some "design for privacy" principles with respect to the RAVE media space environment at EuroPARC, including some specific design suggestions that demonstrate what design for privacy is all about. Langheinrich [136] gives six principles for privacy in ubicomp: notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, and access and recourse. Here, "proximity and locality" means using location information to enforce access rules based on where the accessor is; "access and recourse" means users should be able to access their personal information and see how it has been used by others (i.e., via usage logs).

Lederer et al. [139] promote a so-called "situational faces" metaphor for privacy settings in ubicomp environments. Users, they argue, need adequate *notice* that describes the "situation" to the point where they can *consent* to the appro-

priate "face" (like a user profile; e.g., "secure shopper," "anonymous," "hanging out with friends"). Each "face" could cover multiple situations. Their focus on notice and consent comes from the fair information practices, which also include, e.g., access, security, and redress. The authors also cite the Boundary Principle, which says to place privacy notices[2] at the *boundaries* between different ubicomp environments. The same authors have explored two factors, *inquirer* and *situation*, via a questionnaire-based study [140]. Their study addressed the question, e.g., of whether a person would change "face" if the same inquirer requests access to personal information in two different situations.

Langheinrich [137] and Hong and Landay [104] propose specific architectures for handling privacy issues in ubicomp environments. The architecture in Langheinrich [137] is based on the six principles given by Langheinrich [136], and uses the "machine-readable format for privacy policies on the Web (P3P)" described by Reagle and Cranor [189]. Hong and Landay [104] developed "Confab, a toolkit for building privacy-sensitive ubicomp applications." User needs were assembled from a variety of interviews, papers, and other sources, and then summarized into four categories, including plausible deniability (i.e., ability to give an excuse without the system proving you a liar) and special exceptions in emergency situations—namely, sacrificing privacy for safety.

Closely related to ubiquitous computing is the concept of the Internet of Things (IoT), which focuses more on objects (artifacts) imbued with computing power and

[2]Some researchers are seeking to improve the privacy notices themselves. See Kelley et al. [123] for an example inspired by the nutrition labels required by the U.S. Food and Drug Administration.

network connectivity. Atzori et al. [16] survey the topic and lists a few privacy-protection strategies being considered, including anonymization of data collected by sensor networks and use of privacy brokers between users and services. Weber [252] gives a legal scholar's perspective of security and privacy issues with the IoT, including a short survey of technical requirements for a privacy protection system as well as some existing privacy enhancing technologies. The author also evaluates the role and actions of the European Commissions with respect to security and privacy in the IoT up until late in 2009.

2.3 Why Privacy matters for Human-Robot Interaction

How can a robot threaten somebody's personal privacy? This section examines this question in terms of two human-robot interaction paradigms: autonomous robot behavior and robot-mediated presence. We then consider the question of whether these privacy risks are unique to robotics technologies or shared by similar things like virtual avatars, Internet of Things (IoT) devices, and dolls.

2.3.1 Can Autonomous Robots threaten our Privacy?

Autonomous robots can certainly threaten our *information* privacy inasmuch as they can collect, process, and transmit personal information about us. A robot could also threaten other types of privacy such as territory and solitude, too, although it probably depends on the extent to which it is seen as a *social actor*.

The degree to which a robot is seen as a social actor can vary. HRI researchers have operationalized social actorhood in several different ways, including the social facilitation effect [211], responses to a robot's greeting and other linguistic factors [73], cheating deterrence [103], and the act of keeping a robot's secret [118]. Other HRI studies have also tapped related constructs, such as anthropomorphism [22] and theory of mind [147]. Several such constructs might relate to privacy in HRI, and studying this connection requires the expertise and close involvement of social scientists.

2.3.2 Specific Privacy Concerns raised by Autonomous Robots

Ryan Calo, a law professor, wrote the "Robots and Privacy" chapter of *Robot Ethics* [45]. He identifies three specific privacy risks posed by robots: surveillance, access to living and working spaces, and social impact. The worry about access is worsened by the work by Denning et al. [59] wherein the authors demonstrate security vulnerabilities on several toy robots. Calo divides his concerns about social impact into three parts. First, having robots everywhere could make solitude hard to find. There is some push for this to happen; the South Korean government, for example, officially intended for a robot to be in every household by 2015. Robots can also act as information extractors and persuaders, perhaps better than humans can. Over a decade ago, the ELLEgirlBuddy was deployed to advertise Elle Girl Magazine to teenagers, from whom it also harvested marketing information via instant messaging [45]. Calo also introduces the idea of *settings privacy*, or concern

that the way you personalize your robot could be sensitive information about you; what before was kept to yourself is now *datafied* and stored onboard the robot. This is especially concerning in light of David Levy's 2007 book *Love + Sex with Robots* [149]. As robots become increasingly amenable to customization, and also able to learn from their experiences and interactions, an increasing amount of personal information may become embedded in our robots. In this case, lawmakers should be concerned about the "third-party doctrine," under which individuals can lose Fourth Amendment protection for information they voluntarily give to some third party [45]. As it stands in US law, you may lose your claim over any personal information that your robot learns about you.

Additional robot-specific privacy concerns have been proposed at We Robot, an annual conference about robotics law and policy. Thomasen [240] discusses the possibility of robotic interrogation. Bankston and Stepanovich [19] discuss the interception of email messages by the US National Security Agency (NSA), arguably by web-crawling robots that make decisions about the data. Hartzog [99] points out that robots in particular can be unfair or deceptive to consumers. The issues raised by all these authors demand two responses: study of the real privacy phenomena by HRI researchers and technologies that can be implemented on real robots to help protect user privacy.

2.3.3 Robot-Mediated Presence and Privacy

In the robot-mediated presence paradigm, personification is natural because the robot really is (representing) a person. One simple form of robot-mediated presence is that of Remote Presence Systems (RPSs, already mentioned above). These robots present a video chat interface with the addition of a mobile base so the remote operator can look around and even drive from place to place. One can see how this is very much a remote presence, as one could even participate in a tour or meet-and-greet from afar. Examples of RPSs include the InTouch Telemedicine robots [111], the Beam RPS [229], and the VGo robot [246]. See Lee and Takayama [141] and Beer and Takayama [26] for two evaluations of RPSs in realistic scenarios. Robots could also be mediators, or even full-body surrogates, for persons with severe motor disabilities (see Chen et al. [49]).

It seems clear that remote presence systems, like autonomous robots, cause concerns about privacy. Without the mobile base, RPSs are essentially video media spaces, which have a slew of privacy problems themselves (see Boyle et al. [36] for a review). Moreover, adding the mobile base adds new privacy concerns [20]. RPSs can be driven into private spaces, or used to look around at things against the will of the local user(s). With video chat software like Skype, the local user controls the direction of the webcam; with RPSs, the remote operator has this control.

2.3.4 Are the Privacy Concerns presented by Robots Unique?

Robots share characteristics with some other classes of devices (e.g., personal computers, Internet of Things devices), which raises the question: why do we need to study privacy in human-*robot* interaction as a separate effort? Why not consider it part of IoT privacy, for example? We think robots (which we see as a general category that includes androids, mobile robots, robotic toys, drones, and several other subcategories) have a unique *combination* of properties even though they share each of these properties with at least one other type of device. The one exception is easy to overlook: the label "robot," which has been given particular meanings by different cultures, especially from the influence of science fiction.

Besides being called "robots," robots share their distinguishing, privacy-relevant properties with other types of devices. Robots are controlled by computers, so they can store, process, and transmit data just like computers can. Robots are similar to smart phones, other smart devices, and some wearables and IoT devices inasmuch as they have onboard sensors. Some robots can move from place to place like remote controlled vehicles, although robots are often autonomous to some extent. Robots are also similar to AI agents like chatbots inasmuch as they can interact with users via text, and other robots are like AI personal assistants (e.g., Cortana, Siri, and Alexa) inasmuch as they can use voice. Some artificial agents can even use nonverbal communication via a virtual body, but robots have the extra implications of *physical* bodies [184]. Note that these different modes of communication enable a robot (or similar device) to be attributed a "personality" by users. Fi-

nally, robots can do actions to change things in the real world, a property they share with animatronics (which are not interactive), some IoT devices (which cannot move around the room, although neither can some robots like SPRITE and KeepOn), and non-robotic industrial automation (e.g, a piston that actuates every 10 seconds; these use little or no onboard sensing).

Each of these properties has implications for privacy. Since robots can move around without being carried or worn by a human, they can enter someone's territory without permission. Robots can also switch between taking measurements from a fixed position (e.g., the kitchen) and tracking an object of interest like a person or group of people. Being able to take physical actions includes touching, grasping, and manipulating objects, including personal objects. Robots could also touch people's bodies. Finally, robots can be socially interactive and even affective, so they can *evoke* behaviors from people. For example, a robot could try to use its embodiment, voice, simulated emotions, and so on to build up trust with someone and then extract personal information or evoke embarrassing behaviors.

Our conclusion from this short discussion is that privacy research for human-*robot* interaction should draw from findings about privacy in interactions with related types of devices, but should also be studied separately because robots have a unique *combination* of characteristics. Also, the words people use to talk about different devices can be important, and the word "robot" (as well as words like "android," "drone," and "cyborg") have special meanings.

2.4 Introducing "Privacy-Sensitive Robotics"

2.4.1 Definition

We use the term "privacy-sensitive robotics" to describe work on designing, programming, and using robots in a way that considers users' privacy. This includes work on *understanding* privacy in human-robot interaction as well as on algorithms, design processes, and other techniques for *protecting* it. Privacy-sensitive robotics work can be thought of as a subset of work on human-robot interaction [91]. Although privacy is a social value (i.e., concerning interaction between people) a robot does not need to be intended for social interaction (like the robots in Fong et al. [75]) to be relevant to privacy.

2.4.2 Published Work thus far

A few studies have been published besides ours that we would consider privacy-sensitive robotics research[3]. Syrdal et al. [232] studied disclosure of personal information and Caine et al. [43] studied privacy-enhancing behaviors. Wong and Mulligan [256] studied what concept videos can communicate to potential users about privacy. Denning et al. [59] demonstrated security vulnerabilities in commercially available robots. Lee et al. [142] studied users' perceptions of privacy with a social robot in the workplace. Hubers et al. [108], Butler et al. [40], and

[3]Work must have privacy considerations as at least one of its main goals for us to call it privacy-sensitive robotics work. Mentioning privacy concerns, finding something about privacy in some data, or suggesting privacy protection as an unintended benefit of a new technology doesn't count, although of course we pay close attention to such work.

Klow et al. [127] studied the trade-off between filtering the robot's video feed to protect user privacy and keeping it unfiltered so the remote operator can use the robot effectively. Wagner [249] presents an architecture for automatically detecting private objects and locations. Tonkin et al. [243] studied the effects of embodiment, and Vitale et al. [247] of transparency on privacy judgments. Also, Krupp et al. [132] studied what privacy concerns people have about telepresence robots.

There has also been some theoretical work in privacy-sensitive robotics. Legal scholar Calo's chapter on privacy in *Robot Ethics* [45] identifies three ways that robots present new privacy concerns: direct surveillance, increased access, and social meaning. Another paper by Calo focuses on how drones [44] might prompt changes in U.S. privacy law, and legal scholar Kaminski comes to a similar conclusion in her discussion of home robots and privacy [119]. Kaminski et al. [120] present several categories of potential privacy harms by robots, then some technological as well as legal solutions. Lutz and Tamò call for a new class of jobs to bridge the divide between privacy regulators and engineers [154], and have also argued for the usefulness of actor network theory (ANT) in their analysis of privacy in healthcare robotics [155]. Shulz and Herstad [215] apply the privacy framework by Palen and Dourish to a mobile robot in the home; similarly, Sedenburg, Chuang, and Mulligan [217] apply the Fair Information Practice Principles (FIPPs) as well as research ethics to the development of therapeutic robots in the home. Finally, Ishii [112] discusses legal frameworks for privacy regulation in the EU, US, Canada, and Japan with a focus on Privacy by Design (PbD).

We now move to an outline of *our* contributions to this area of research.

2.5 Preview of Contributions

When we began researching what we now call "privacy-sensitive robotics" we didn't even know the extents of the topic—i.e., what all is meant by the word "privacy" and how it could relate to human-robot interaction. This motivated an ambitious literature review. As we were reading privacy literature we also became curious about how well-equipped society is to address privacy concerns (assuming we knew what they were) in terms of the needed technology. The findings from this extra search form our first contribution in this dissertation (Chapter 3). Specifically, we searched for technologies that could provide *constraints* on the perception, navigation, or manipulation of a robot to protect users' privacy. We found a few technologies that were designed for privacy protection (Section 3.2) and many more that could potentially be used for privacy protection (Section 3.3).

Our next three contributions are findings from empirical studies. Our first concern was *visual* privacy—i.e., objects, areas, and people that someone doesn't want others to *see*. We defined an objective metric for visual privacy: are all the objects, areas, and people that have been marked as "private" completely obscured (e.g., blurred) in the robot's video feed at all times? This approach works well if (1) some objects, areas, or people are inherently private, like tax forms or naked bodies—simply detect and obscure them automatically; (2) users specify their preferences manually for every object, area, and person; or (3) a compromise between options 1 and 2 wherein users specify their preferences over *categories* of objects and an automatic classifier decides which objects in the video feed belong to which

of these categories. Following option 2, our first experimental contribution (Chapter 4) evaluates different interfaces for tagging objects as private—in particular, whether physical interfaces make this process easier than on-screen interfaces do.

As we finished this first study, we began to think about what might cause privacy preferences to *change*. In many cases, for example, objects are private not because of what they are in themselves, but because of a person's unique relationship to that object (e.g., objects with sentimental value). Taking this a step further, sometimes privateness is not fully explained even by the person-object relationship—in fact, Nissenbaum's theory of "contextual integrity" [172] stresses that privacy rules change based on the situation.

Following these thoughts, we argued in our second experimental contribution (Chapter 5) that contextual "frames" are a major influence on how people feel about their privacy. We conducted four online surveys to look for differences in how people interpreted a scene when the "frame" was changed. In the final survey participants watched four video clips of different parts of a human-robot interaction scenario, and we began to test for any effects of their adjusting to the robot's presence over time. Although these initial results hinted that time was indeed having an effect, we wanted to look at this over longer periods of time and in a more natural setting.

Pursuant to this new goal, our third study lasted six weeks and happened in a public place: a hallway on campus outside a yoga classroom. Instead of measuring privacy concerns directly we became interested in how people form a mental model of how the robot works, which we see as one of the main inputs to the

process of making privacy judgments. We interviewed six participants repeatedly throughout the study to learn about mental model formation. Our fourth contribution (Chapter 6) is a report of the key findings and recommendations from those interviews.

We think of these early, pioneering contributions we have made to privacy-sensitive robotics as a journey or story. We first chose a broad definition of privacy and found the state of the art in privacy protection for robots. We then chose for our first study a rather narrow problem with a practical outcome—how should users tag objects as private?—and then expanded our focus in the second and third studies by removing two big assumptions: what if the situation is framed differently? and, how about when users' mental models of the robot change as they interact with it? Because of our desire to get other researchers involved in privacy-sensitive robotics work, our fifth and final contribution is a set of recommendations for the future of research in this emerging area. We hope that the collaborations we describe and the roadmap we lay down will guide and inspire the choices made by a new wave of privacy-sensitive robotics researchers.

3 Perceptual and Behavioral Constraints that could Protect Users' Privacy

This chapter was adapted from a paper by Rueben and Smart [195] that was presented at We Robot 2016.

3.1 Introduction

Roboticists usually conceive of robot behavior as a list of "do"s: do pick up the cup, do go to the other room, do locate the red object. Along with each "do," or *goal*, however, comes a host of "do not"s, or *constraints*. Do not drop or crush the cup; do not hit anyone on your way to the door; do not stare at people or they'll feel uncomfortable. In many robotics applications, the proper constraints are apparent from the outset, or they become apparent as a natural part the design process. They can be hard to ignore—e.g., for an urban delivery robot it is obvious and urgent that it should stay on the sidewalk to avoid becoming the victim of a traffic collision. It should also be obvious to designers of, for example, a robotic receptionist for a museum or hospital that it should not interrupt or speak rudely to users, and that movement should be constrained if someone gets too close.

Some constraints, however, require special attention to ensure they are not forgotten or ignored. For example, constraints to uphold values like safety, ac-

cessibility, etiquette (e.g., Takayama et al. [235]), transparency[1], and our value of interest: privacy. There are several reasons why the proper constraints to protect user privacy might be left out of a robotic system. First of all, they might not be so obvious to designers—e.g., it might be hard to predict when a social robot might cause subtle harms to someone's close relationships. Designers might also not be motivated or encouraged to address certain privacy concerns, so they simply ignore them. For example, a security robot might pass too close to people or stare at them creepily, but the building owners who hired the robot service might not care. Lastly, robot manufacturers could even make money or gain an advantage by exploiting people's privacy, e.g., by harvesting people's data for targeted advertising. These problems may seem minor or speculative now, but we believe they will become much more serious as robotics technology improves and becomes more ubiquitous.

Our first contribution to privacy-sensitive robotics research addresses these issues by focusing on how to *constrain* robots so they respect people's privacy. In particular, we have surveyed the technology literature and assembled a list of existing technologies that could be used to implement the proper constraints. We also present some work that has already suggested or implemented privacy constraints for robots. The purpose of this first contribution is to understand what is the state of the art in privacy protection technology for robots.

[1]By *transparency* we mean that the robot's appearance, behaviors, and interface make it clear what is going on inside the robot.

3.2 Constraints for Privacy-Sensitive Robots

We will need to restrict what the robot sees (Section 3.2.1), where it goes (Section 3.2.2), and also what it touches (Section 3.2.3). What a robot (or robot operator) actually *does* with the data it collects or in the spaces it occupies is not being considered here, although data usage and robot behavior will impact users' privacy expectations.

It should be remembered that constraining a robot can draw extra attention to private objects or areas. If a remote operator can see that an object is being redacted in a video stream, he or she might wonder what is being hidden; if the robot conspicuously avoids an area or object, local users might become curious. If a remote operator is malicious, or even just curious, he or she might move the robot to try to see past a filter or manipulate a filtered object. Perhaps additional, fake restrictions can be added to make the real ones less interesting by comparison. In any case, good privacy-protecting robot constraints should ultimately be tested against malicious users to ensure they are robust to tampering. If a constraint that is perceivable by remote operators or local users would attract too much attention, that constraint might have to be made imperceivable or removed altogether.

3.2.1 Constraining Perception

Many robots are equipped with cameras. Here we will consider how to limit what a robot can see. Some robots can hear, feel, or even smell, but it is more difficult for these senses to violate someone's privacy than it is for vision. We have said that

privacy can be informational (which includes seeing and hearing private things) as well as spatial. Some of the most basic privacy issues are visual: seeing someone without their clothes, for example. Also, many informational privacy violations among humans are visual: intellectual property is stolen by reading a document or seeing a product. Even spatial privacy may have a significant visual portion, as indicated by the trope of blindfolding outsiders as they are led through a secret area.

It is important to note from the outset that image manipulation is only important if the vision system actually captures privacy-compromising data. For example, Zhang et al. [262] present a computer vision system to detect if an older adult has fallen down. Their work points out that certain sensors—here, a depth camera—can function without compromising a person's identity. They also use an RGB camera, which, by collecting color image data, affords much less privacy protection. It becomes clear that, when a user is beyond the range of the depth camera, a decision needs to be made as to whether to use the RGB camera, which compromises the user's identity, to continue providing functionality. Once it is decided that privacy-compromising data will be collected, image manipulation techniques may be employed to mitigate the privacy violation.

Templeman et al. [238] present a classifier for deciding whether an image was taken in a private area such as a bathroom or bedroom. The classifier attempts to match landmarks in the image stream with landmarks in sample images of the private areas. The authors focus on first-person cameras for humans, but their work probably extends to cameras onboard robots. This becomes unnecessary for

robots that can localize themselves with high confidence in a map with the private areas labeled; the classifier remains useful for robots that do not localize or do so with too much uncertainty.

Forsyth et al. [76] present a technique that detects whether one or more naked people are in an image and, if so, provides a mask of the offending region. Privacy-sensitive robots could use a nudity detector to *avoid* naked humans in some contexts (e.g., upon opening a bedroom door and seeing the occupant in the middle of changing his or her clothes, the robot could decide to leave immediately) and, in other (e.g., medical) contexts, to *cover* the naked bodies with an image filter (see next section for examples) before the images reach the remote operator. It seems that most people regard certain of their body parts to be private—i.e., off-limits for viewing by others—when uncovered. A nudity detector would help robots respect this aspect of privacy. It appears that the state of the art has a long way to go if robots are to reliably respect this form of privacy. Difficult unsolved problems include automatically deciding whether a partially-dressed person is "decent" (i.e., has all the private body parts covered) and differentiating between, e.g., a swimsuit, underwear, and lingerie, each of which could evoke a different set of privacy rules. Context also matters—e.g., when a woman is breastfeeding an infant it might be acceptable to expose more than usual.

3.2.1.1 Image Manipulation Techniques: Descriptions.

Many techniques can be found in the graphics and animation literature for post-processing images. Note that these techniques only use the information available within the image itself; outside help from the artist or from additional (e.g., depth) sensors is not being considered. Some image manipulations are intended to obscure parts of the image. These methods include pixelating, blurring, redacting, and replacing either the entire image or just certain regions (see pictures in Boyle et al. [35], Hubers et al. [109], Raval et al. [188], and Zhao and Stasko [263]). Only a few works are discussed here, as representative examples.

Perhaps the simplest way to simplify a digital image is to reduce its resolution. This makes the image appear blocky, hence the moniker, "pixelation." Naive pixelation preserves low-detail regions but obscures high-detail regions of an image. For example, in Figure 3.1b the subject's eyes, nose, and mouth are difficult to discern, but his clothing remains discernable as a suit and tie. With more intelligent pixel grouping, it becomes easier to discern the eyes, nose, and mouth, and arguably without making the subject's identity more recognizable (Figure 3.1e).

Blurring and redaction are also common methods for obscuring images, especially on post-processed television. Blurring smooths an image by allowing each pixel's value to be influenced by the values of the pixels around it. Including a larger neighborhood ("kernel") of pixels around each target pixel makes the image blurrier. Redaction is simply removing pixels from an image, yielding the familiar black box over the objectionable part of the image. Redaction benefits from an

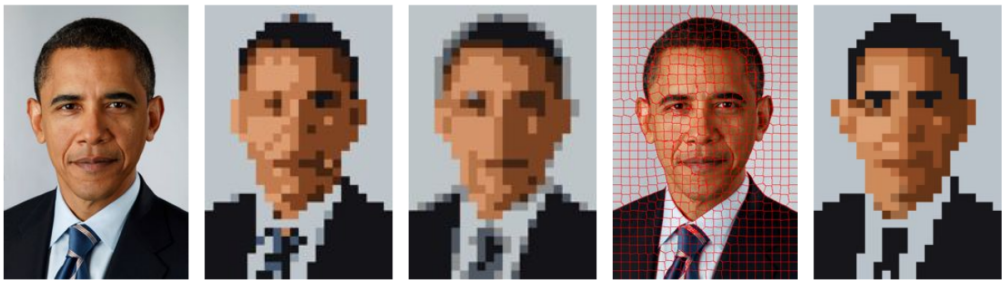

Figure 3.1: Reproduction of Gerstner et al. [85, Figure 1]. Original caption reads: *Pixel art images simultaneously use very few pixels and a tiny color palette. Attempts to represent image (a) using only 22 x 32 pixels and 8 colors using (b) nearest-neighbor or (c) cubic downsampling (both followed by median cut color quantization), result in detail loss and blurriness. We optimize over a set of superpixels (d) and an associated color palette to produce output (e) in the style of pixel art.*

intelligent way to select objects in an image; one such method is GrabCut [193], which uses a partial classification of foreground and background pixels to intelligently divide the image in two along the object boundary. GrabCut has recently been extended to use depth as well as color information [245].

Korshunov and Ebrahimi [130] show the possibility of morphing images of faces so they are unrecognizable and don't need to be blurred or redacted altogether. Building on that initial work, the method proposed by Nakashima et al. [166] preserves facial expressions and is shown to work on a few realistic images.

Replacement is like redaction, but uses something more purposeful than a black box to cover the redacted areas. A familiar example is the use of a chroma key or "green screen" technique to replace the background of a movie set or news studio with a computer-generated, moving environment. If we know what sort of thing

has been selected, we can do specialized replacements. For example, if a person detector returns the positions in the image of a person's joints, we could cover the person with a generalized cartoon in the same pose. Replacing an object with what "would be behind it" as described by Hubers et al. [109] requires knowledge that a camera cannot give, since the object occludes that area. One way to circumvent this is to use the other information in an image to paint over the object such that it's hard to tell that the image has been altered. This is the strategy of so-called "inpainting" or "image completion" techniques like those by Sun et al. [230], Cheung et al. [51], Bugeau et al. [38], and Herling & Broll [102]. Cheung et al. [50] apply inpainting to surveillance video. The key idea here is that, in order to perform inpainting, one must somehow know what is behind the target object without actually seeing it. For certain scenarios, this will be impossible without more information.

That additional information could be recorded beforehand if the environment does not change much over time. Recently, sensors that fuse color and depth information have made it possible to build colored 3d maps; using a mapping framework like OctoMap [105] with simultaneous localization and mapping as in RGB-D SLAM (demonstrated by Endres et al. [64]) could very well provide the knowledge for intelligent image replacement in semi-static environments.

Whereas our goal thus far has been to review methods for *obscuring* objects in images, most graphic artists have different goals in mind. They seek to make images simpler and less busy [134], or more attractive [153], or easier to understand [56, 57]. Nevertheless, the techniques they present could also be useful for protecting user

privacy in robotic systems.

Such techniques are typically categorized as non-photorealistic rendering (NPR), as opposed to techniques that focus on fidelity to photorealism. So-called "painterly" techniques (e.g., by Lu et al. [153]) use a variety of different brush-stroke effects to make images look more artistic, incidentally removing identifying details in the process. Image abstraction techniques seek to retain the gist of an image and discard the distracting details. Kyprianidis [134] uses the anisotropic Kuwahara filter, which aims to provide a consistent level of abstraction across different levels of detail while also being robust to even higher-contrast noise. Human faces are still quite recognizable (as intended) after applying this filter. DeCarlo and Santella [56] add definition to an abstracted image by introducing certain of the detected edges that were removed during abstraction. The particular implementation by DeCarlo and Santella is not so practical for our purposes because it requires users to designate which areas need more definition. In fact, these areas are designated implicitly via eye tracking. One cannot help but wonder: would it be effective to do the opposite by starting with all the edges and *subtracting* the important ones in order to obscure features of interest? Conversely, perhaps removing color (and, therefore, textures) could also protect privacy. DeCarlo et al. [57] present a way to represent 3d models of objects using only the surface contours and some "suggestive" extensions of those contours. The technique is automatic and can yield recognizable renderings of the models, but extending it to 2d images without depth information might yield results that are less satisfactory.

Artists, like magicians, are masters of attention control. They possess an arse-

(a) Emphasis on left (b) Emphasis on right

Figure 3.2: Reproduction of Lu et al. [153, Figure 8]. Original caption reads: *Placing emphasis via controlled stroke density.*

nal of ways to make the viewer focus on something, or (since we are talking about privacy concerns) to *not* focus on something else. Cole et al. [52] present several methods for subtly manipulating which regions are emphasized in 3d drawings. This could be thought of as indirect obscuration, as the regions *not* emphasized experience dramatic reductions of detail. Of course, if obscuration is the goal, any user control over the "gaze direction" would have to be limited. Another compelling illustration (no pun intended) of emphasis control—here, by controlling stroke density in a painterly filter—is shown in Figure 3.2 (reproduced from Lu et al. [153]).

3.2.1.2 Image Manipulation Techniques: Evaluations.

From the standpoint of privacy, the primary goal of image manipulation is to obscure private objects. Depending on the application, the secondary goal could be either to maximize the fidelity of the rest of the image (e.g., so the user can use

the image feed for some purpose or task) or to hide the fact that manipulation has occurred, or both. From the standpoint of non-photorealistic rendering, however, the primary goal of image manipulation is to enhance or simplify the image; privacy is not necessarily being considered there, and the goals may conflict. Here we review some evaluations of how image manipulation impacts image perception by humans.

We frame this section with the model of (visual) privacy loss proposed by Saini et al. [207] for multi-camera surveillance systems. They define privacy loss as the product of "identity leakage" (i.e., the probability of being identified) and a "sensitivity index" (i.e., how sensitive the information is that you're being identified with). Their model considers not just facial recognition, but also the "what" (clothes, gait, behavior), "when" (daily schedule), and "where" (location) as additional inference channels for identifying a person. The authors also discuss some privacy implications of *who* exactly is seen by a multi-camera surveillance system; e.g., if the same three people are always visible, it becomes easier to discern their identities because there are fewer possible identities to choose from.

Early evaluations of image manipulation to protect privacy are on simple filters in video media space applications. Zhao and Stasko [263] present five filtering techniques and evaluate them on short video segments in a privacy context. The filters were pixelization, edge detection, and three other techniques that were basically abstractions. All filters appeared to make the actors more difficult to identify without making it much more difficult to discern what they were doing. Subjects made an interesting comment: actors may be easily identified by shirt color even

after heavy image filtering.

Boyle et al. [35] are also concerned with filtering short video clips for privacy. Two filters—blur and pixelize—are each used at 9 different fidelity levels. It is shown that, for each filter type, there was an appropriate fidelity level for which privacy was protected and some basic awareness information (e.g., number of people in the room) was preserved.

Schiff et al. [213] focus on reliable tracking of color markers that people wear in order to have their faces obscured in a video feed. They test their system at a construction site using high-visibility vests and hard hats for markers. The robustness of their system to lighting conditions and partial occlusions is examined.

Korshunov et al. [131] present a crowdsourcing approach to evaluating privacy filters using Facebook. These online results are validated against an in-person, laboratory study. The study itself shows that privacy is most preserved by a redaction filter, followed by a pixelation, and least of all by a blur filter.

Halper et al. [96] provide our introduction to the psychology of how people perceive NPR images. First, separating foreground from background is facilitated if everything in the foreground is rendered with one style of NPR and everything in the background is rendered in another. Perhaps private objects could be de-emphasized by rendering them like the background, and key objects could be intentionally emphasized by the reverse process. Second, people tend to perceive sketchy renderings as more open to change, and want to explore areas with a higher level of detail. These findings appear useful in mapping, both for visualizing (un)certainty about different regions and for guiding a user to key places (and away

from irrelevant places, thereby protecting some privacy). Influencing the viewer's attention might only protect privacy for cursory viewings, though; stronger image filters might be needed to protect against ill-intentioned viewers or viewers who spend more time scrutinizing the scene.

If we use NPR to remove private details or replace private objects seen by a remote presence system (RPS), will users still be able to complete their tasks? That is, will users still accurately interpret the scene? It is crucial to see the difference between artistic renderings, in which different users are expected to interpret the product in different ways, and functional renderings such as are found in technical manuals, which seek to communicate the same things to everyone [90]. NPR provides a "functional realism" that is well-suited for the latter purpose, but several studies have shown how non-photorealistic environments can distort user perceptions of a scene. Gooch & Willemsen [90] demonstrate that people typically underestimate distances in NPR immersive environments, e.g., while walking down a hallway. In a more extensive study, Phillips et al. [183] compare a NPR virtual environment to a highly realistic one and confirms that a lack of photorealism in particular seems to cause the distance judgment error. The authors draw on earlier findings to make a non-intuitive point: it wasn't the *reduction of detail* that caused the misjudgments, but (so it appears) the *reduction of photorealism* itself. It seems, then, that using NPR to protect privacy could inhibit tasks for which distance judgments matter.

Perhaps the most important consideration for using NPR to protect privacy is how it affects the *discernability* of objects; that is, whether an object is there,

what sort of object it is, and whether it's real or virtual. Fischer et al. [71] address the last of these questions. The authors test a promising strategy: if an image containing both real and virtual objects is uniformly stylized, does it become harder to discern between them? Yes, the results showed that users had more trouble discerning virtual objects from real ones with this "stylized AR" technique. This is promising both for immersive, realistic AR and for convincing object replacement in a privacy protection system.

Several papers directly address using NPR to protect privacy. Erdelyi et al. [65] compare a cartooning (i.e., abstraction) filter to blur and pixelation filters in terms of privacy protection and utility. Abstraction provided the most utility and the least privacy, followed by blurring, then pixelation. The same authors present a modified version of their technique [66] as an entry into the MediaEval 2014 Visual Privacy Task (see Badii et al. [18]). Images containing people were abstracted once, then additional abstraction and pixelation filters were applied to just the faces. The filter performance was evaluated by crowdsourced survey responses. Output images were rated in terms of intelligibility, privacy, and pleasantness, and scored near the median performance for the competition.

3.2.2 Constraining Navigation

Robot navigation must be constrained if we are to prohibit the robot from entering certain areas. For example, bedrooms or children's play areas might be private in many circumstances and therefore off-limits for mobile robots. Several methods

for enforcing these constraints are discussed below.

3.2.2.1 Motion Planning and Obstacles.

Motion planning algorithms typically use a model of the obstacles in the world to restrict the valid planning space for the robot. In this framework, the "configuration space" includes all valid configurations of the robot such that it is not in collision with or too close to an obstacle (see Ch. 4 of LaValle [138]). This can be readily adapted to include areas that are private in the sense that the robot must not enter or pass through them; those areas can simply be designated as virtual obstacles. This will have predictable effects on the plans produced by the chosen planning algorithm. The new privacy "obstacles" blot out part of the search space and paths become either (a) less optimal, (b) invalid, or (c) the same as they would be otherwise. If the search space includes a temporal dimension, we can also model obstacles that toggle on and off or change size over time. For example, the bedroom may be off-limits only at night, or a morning person might desire more personal space in the early hours of the day. Besides the added dimension, the planning problem stays the same conceptually with the addition of time information; that is, as long as the time spent planning is trivial.

What if these private regions are not defined beforehand, or what if they change periodically? Here we can use algorithms for mapping and planning in an unknown environment. When the obstructed space changes (i.e., when a private region is turned on, turned off, or moved), the robot will have to use techniques summa-

rized by Russell & Norvig [201], Ch. 12.5–6, such as plan monitoring, replanning, and continuous planning. Private regions that move around, such as the personal spaces of people, might warrant the use of "differential" planning constraints (see Ch. 13 of LaValle [138]), which regulate velocity and acceleration in addition to position. Obeying these constraints becomes more difficult for more complex robots, and frameworks such as the "whole-body control framework" for humanoid robots presented by Sentis & Khatib [218] become necessary for planning natural, stable motions.

3.2.2.2 Semantic Maps.

The maps used for constraining navigation in the above discussion were metric maps. Metric maps are to-scale, cartesian representations of a space. Time could be added as an additional dimension, but the map remains purely quantitative. The problem is, humans don't typically think about their environments in terms of numbers. With the exception of temporal information (e.g., "between the hours of 9 and 11 A.M."), people typically use qualitative labels to refer to objects, not spatial coordinates. For example, "the document on my desk" is used in lieu of, "the 8-1/2in x 11in rectangular object of nominal RGB color value (113, 7, 24) located at (x,y,z) = (1.2m, -0.2m, 0.8m) from the. . . ." This suggests the utility of adding higher-level conceptual information to metric maps to form *semantic*, or meaning-laden, maps. The robot could store meaningful labels for persons, places, and things in the world. These labels could then be compared, grouped

into hierarchical categories, and used to inform a robot's decisions.

Galindo et al. [83] present a semantic mapping framework. In the authors' language, conceptual entities in the conceptual part of the map are "anchored" to locations, areas, or objects in the spatial part of the map. Some semantic labels can be assigned automatically, such as grouping spaces into "rooms" and "corridors" based on connectivity. The system can then make inferences such as, "this room with a bed in it must be a bedroom." This framework makes robot navigation easier by allowing commands like, "go to the kitchen." It also makes robot localization faster with inferences like, "I see a TV set, so I must be somewhere in the living room." Galindo et al. [84] show this same framework used in a series of robot task planning experiments. The semantic map framework allows the robot to make inferences from the state information and to shrink planning problems by raising the level of abstraction. Of special note is the robot's ability to intelligently handle commands that are too specific; e.g., when told to go to the bedroom, the robot reasons that all the rooms it has found are candidate bedrooms and proceeds to search for discriminating evidence. This ability to generalize commands when unsure about their meaning is promising for privacy applications.

Semantic maps may be created without manual labeling by a human. Rusu et al. [202] present a system for creating semantic maps automatically from 3d point clouds. That same system is implemented autonomously in a kitchen environment with a PR2 robot by Goron et al. [93]. The robot is able to explore, generate a hypothesis map, and then interact with appliances to verify and revise the map. These semantic mapping frameworks are promising for embedding privacy settings

in the robot's model of the world, which can then be honored by avoiding certain actions.

3.2.2.3 Rules for Moving amongst Humans and other Robots.

Privacy can be thought of as including personal space, since violating one's personal space violates both one's solitude and one's control over access to oneself. The robotics community has addressed personal space in mobile robot navigation via the notion of proxemics, introduced by Hall [95] almost sixty years ago. Since proxemics was originally defined as being between humans, a new array of studies is needed to discover any differences in the human-robot interaction scenario. Such work exists; for example, Butler & Agah [41] studied what types of approach behaviors make humans uncomfortable, and Takayama & Pantofaru [234] included some human traits as factors when they studied this in more detail. Findings from such work will inform the ways we constrain robots to respect personal space. Several robot behaviors have already been implemented with personal space in mind: standing in line [167], following a person [89], and passing a person in the hall [151, 152]. Some other relevant but older studies include that by Asama et al. [15] for robot-robot passing behaviors and another by Kato et al. [122] for handling human-robot traffic through passageways and workspaces.

When a robot passes a person using standard approaches to navigation, it often violates personal space by passing too close, possibly without slowing down. Lu and Smart [151] change the standard costmap implementation in several ways (further

techniques are explained by Lu et al. [152]) to incentivize the path planner to give the human more space. Also, constraining the robot's gaze direction matters in this scenario. Lu and Smart predict that the robot needs to look at the human at least once to acknowledge that his presence is accounted for, but a constant stare is "creepy." The user study concluded that the costmap manipulation increased passing speed for humans, but that an intermittent eye contact policy did not; gaze effects seem to be more complicated than they thought.

3.2.3 Constraining Manipulation

We have said that being sensitive of privacy restricts what the robot should see, where it should go, and also what it should touch. We now turn to that last element, perhaps the least-explored of the three: how can we restrict the robot from *touching* things in ways that would violate privacy? This could include touching personal possessions, objects in a person's territory, or even a person's body. Also, *touching* is too narrow here; some things may be inappropriate to pick up but OK to touch, whereas other things might be inappropriate even to reach towards or point to. All these actions fit within robotic manipulation, or at least motion planning for robot arms.

During autonomous operation, applying privacy constraints to robotic manipulation could mostly be handled either by modeling private objects as obstacles (see Section 3.2.2.1) or by labeling them appropriately (e.g., "NoTouch" or "NoGrab"; see Section 3.2.2.2). In the first case, a typical trajectory planner would

plan around the private object. In the second case, a high-level, semantic planner would reject plans that include illegal actions (e.g., "Touch the Book that has label NoTouch").

During teleoperation, an obstacle-based approach could still be used to simply take away the user's control whenever some action would violate privacy. This does not, however, give much feedback to the user. Perhaps an improved method would use a haptic device to deliver force feedback to the user when the robot nears a restricted area. The privacy boundary could exert an increasing amount of normal force via the haptic device as the user nears the boundary's edge. Rydén presents a relevant framework called "forbidden-region virtual fixtures" in his doctoral dissertation [205]. His work enables remote touching of moving [203] objects that considers both the position and orientation of the remote toucher [204]—in our case, a robotic end effector. Perhaps this system could be used in assistive robotics, where humanlike robots might need to touch and move patients while dynamically maintaining appropriate hand positions. It should only be used, however, when it is acceptable for the teleoperator to know about the presence and location of private regions, as is probably the case for private body parts—otherwise, it could be used maliciously to find the things people want to protect!

3.3 Existing Work on Constraints for Privacy-Sensitive Robots

We also found some work on constraints that was already specifically applied to robots and focused on protecting users' privacy. Although we could not find very

much that has been published so far, we present three categories of early work below.

3.3.1 Visual Privacy

Some work has focused on implementing and evaluating tools for protecting visual privacy by obscuring things in the robot's video feed. An initial concern has been the tradeoff between visual privacy and *utility*, or usefulness of an interface for its intended purpose. Jana et al. [114] present a privacy tool for simplifying videos to only the necessary information. Although no user study is conducted, the tool is shown to preserve utility and protect privacy through a simple analysis. Raval et al. [188] present two more video privacy tools. One uses markers to specify private areas, and the other uses hand gestures for the same purpose. Butler et al. [40] coin the term "privacy-utility tradeoff" and test a pick-and-place task with the PR2 robot; Hubers et al. [109] did similar tests for a patrol surveillance task with the Turtlebot 2 robot. Both studies found it feasible to complete the tasks with effective privacy filters in place.

Rueben et al. [198] report on a user study comparing three different interfaces for specifying visual privacy preferences to a robot. This study is presented as our second contribution in the next chapter (Chapter 4).

3.3.2 Proxemics

We have already referenced several works in Section 3.2.2.3 that constrain the robot's movements to respect people's personal space (e.g., Lu and Smart [151]). There have also been several proxemics *studies* in human-robot interaction. These count as privacy-sensitive robotics when the goal is to understand how robots can move and position themselves to promote human comfort and acceptance—positioning optimally for performing a cooperative task with a human does not count by itself. Publications include work by Mumm and Mutlu [165] on both physical and psychological distance, as well as by Okita et al. [176], Joosse et al. [117], and Henkel et al. [101].

3.3.3 Territoriality

Territoriality is a construct of great interest in the social sciences (e.g., in Altman [10], discussed above in Section 2.2.3.1), but has been studied very little in the HRI domain. One work, by Satake et al. [208], develops and tests a model of territory in front of shops. The authors show that people preferred chatting with a robot that understands which territory belongs to the store. We believe that much more exploration of territoriality is warranted in HRI.

3.4 Summary and Areas for Future Study

In general, we have seen that technologies exist for constraining robot perception, navigation, and manipulation that could be used to implement privacy protections on robots. If you want to prevent your robot from seeing something, driving somewhere, or touching something some tools exist that you could start with. Image manipulation techniques seem especially well-developed and diverse.

We have also seen, however, that our focus on "constraining perception, navigation, and manipulation" left out some important categories of technology that will be needed for building privacy-sensitive robots. We conclude this contribution by recommending that technologies that could help in these areas be identified or, if they do not exist, invented. We propose five such areas:

1. **User interfaces** will be needed for specifying users' privacy preferences to robots. Preferences might be vague, abstract (e.g., "don't bother me when I'm thinking"), or complicated. Preferences will also refer to objects, contexts (e.g., don't collect any pictures "when I haven't done my hair yet"), rooms, and times. People might need to communicate any or all of these types of preferences to a robot via its user interfaces.

2. **Cognitive frameworks** will be needed for reasoning about these privacy preferences once they are communicated to the robot. We will need ways to represent privacy preferences, prioritize them, decide between conflicting preferences, and pay attention to contextual factors like the time of day, situation (e.g., whether two friends are fighting or joking), or location (e.g.,

room in a house).

3. Privacy-sensitive robots might also need to **recognize or at least guess what's private**—including situations, objects, conversations, embarrassing states, and rooms—without users specifying it beforehand. A mixture of prior knowledge from the robot's programmers and learning from experience might be necessary for a robot to even begin understanding such an abstract concept.

4. **Privacy filters for non-visual sensor data** will also be important. This will include sound, depth (e.g., from a LIDAR or Microsoft Kinect sensor), and touch (the robot can touch people and objects with its end effectors, and can also feel when people touch it if it has sensors in its skin). Filters might also be necessary for simpler sensors like wheel encoders if they are used to create more abstract data products—for example, the average distance a robotic vacuum has to drive before hitting a wall could be used as a proxy for the size and therefore the value of a house.

5. Tools for constraining robots to respect the different types of **psychological and social** privacy presented in our taxonomy (see Section 2.1.3). Some of the proxemics technologies that we have already discussed in this chapter might be useful here, but we will also need to control gaze behavior, which conversations the robot is part of, and other social behaviors.

3.5 Choosing our Next Contribution

The next three contributions are all findings from *empirical* studies relevant to privacy-sensitive robotics. We present findings from a variety of areas, including interfaces for specifying privacy preferences (Chapter 4), the factors that influence people's privacy concerns (Chapter 5), and how people come to understand the robot's behaviors and sensing capabilities (Chapter 6).

Our next (second) contribution focuses on visual privacy—we assume there are things that the user does not want to be visible in the robot's camera feed. This could include objects like personal electronics or toys or furniture; areas like certain rooms or the space around a person; and people, especially people's faces. We also assume that people have a set of preferences that specify for each object, area, or person what type of filtering would be appropriate to address their privacy concern. Levels of filtering could be as heavy as turning off the camera altogether or as light as a mild blurring effect, or things like abstracting color or texture from images and displaying a diminished image like a map or the outlines of obstacles. Given these assumptions, we proceed with a broad research question: how can a person's privacy preferences best be communicated to a robot?

4 Evaluation of Physical Marker Interfaces for Protecting Visual Privacy from Mobile Robots

This chapter includes work by Rueben, Bernieri, Grimm, and Smart [198] published in the Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2016).

Abstract— We present a study that examines the efficiency and usability of three different interfaces for specifying which objects should be kept private (i.e., not visible) in an office environment. Our study context is a robot "janitor" system that has the ability to blur out specified objects from its video feed. One interface is a traditional point-and-click GUI on a computer monitor, while the other two operate in the real, physical space: users either place markers on the objects to indicate privacy or use a wand tool to point at them. We compare the interfaces using both self-report (e.g., surveys) and behavioral measures.

Our results showed that (1) the graphical interface performed better both in terms of time and usability, and (2) using persistent markers increased the participants' ability to recall what they tagged. Choosing the right interface appears to depend on the application scenario. We also summarize feedback from the participants for improving interfaces that specify visual privacy preferences.

4.1 Introduction

As mobile robots and remote presence systems become more common in our daily lives, we must address the privacy concerns of the people that share a physical space with these systems. In particular, *visual* privacy becomes a concern because people are worried about what is being viewed or recorded. Users may want to control what the robot (or its remote operator) sees, much as they would put away or hide valuables or pictures before a stranger visits. This paper presents a user study that explores three different methods users might use to specify what they want hidden. The context for the study is a naive user's first experience with an autonomous robot janitor that works in an office setting. Since we are working with an actual physical space and real objects two of our methods are *physical*, i.e., they involve pointing at, or placing markers in, the physical environment. We compare these two physical methods with a traditional GUI interface using efficiency, usability, and task performance metrics.

4.2 Related Work

Research has revealed that humans often interact socially with machines. This phenomenon is often stated as *"Computers Are Social Actors"* (CASA) [168]. Any robot, then, can function as a social actor during a human-robot interaction. Broad discussions of privacy issues that are specific to robotics are only recently beginning to be published, especially outside of the robotics discipline. Calo [45] gives a good overview as well as some newer insights.

Privacy is important in all human cultures [11], although different cultures have different norms for privacy and different mechanisms for enforcing those norms. If robots can function as social actors in whichever human culture they inhabit, we want to study how we can enculturate robots with respect to our privacy norms. We call research that studies these questions "privacy-sensitive robotics."

Studies of privacy in video media spaces are relevant to robotics, especially teleoperated robots and remote presence systems (RPSs) [126, 36]. Several studies consider using filters such as pixelation, blurring, and other techniques to retain some scene information but hide private details [263, 35, 85]. Moving beyond just video media spaces, Jana et al. [114] filter images using edge and motion detectors in several different contexts. Raval et al. [188] address the problem of *specifying user-specific privacy preferences* using gestures and a special border for use on flat surfaces. This paper explores that same *specification* problem in a human-robot interaction via a detailed user study.

Privacy is rarely the main focus of human-robot interaction studies. Evaluations of remote presence systems for elderly users have revealed some privacy concerns [33, 26]. Denning et al. [59] have raised concerns that commercially-available robots could be hacked by malicious persons and create new privacy risks for users.

Two early studies in privacy-sensitive robotics are concerned with the *privacy-utility tradeoff*, i.e., how much task performance is sacrificed for a given increase in privacy protection [40, 109]. The goal is to preserve utility while protecting privacy. These studies also show that privacy preferences differ between individuals [37]

and are complicated, so privacy-sensitive HRI research needs to study real people to be useful. It can also be important to study user behavior rather than just asking people questions (e.g., in surveys), as people have been shown to self-report wrongly in certain situations [171]. Along these lines, this work focuses less on privacy protection technology *per se* and more on the interaction between that technology and real end users.

Qualitative user feedback from this study has already been reported, including written responses to survey questions as well as conversations between the experimenter and each participant about interface usability [197].

4.3 Research Questions

Our focus in this contribution is on the comparison of a traditional GUI interface to two different physical interfaces in the context of specifying that certain objects should be "private." Fundamentally, we are interested in the effect of (1) leaving a persistent tag on the object and (2) physically moving close to the object to tag it. We now briefly describe the three interfaces (details are in Section 4.4) that operationalize these properties and then present and justify hypotheses about their relative performance on five different measures.

Our three interfaces are (1) **The Marker Interface:** the user physically places a persistent marker directly on objects; (2) **The Pointing Interface:** the user physically touches objects with a "wand" tool; and (3) **The Graphical Interface:** a traditional on-screen graphical interface where the user clicks on objects.

Interfaces (1) and (2) are collectively referred to as the *physical* interfaces. In all three cases a monitor situated next to the objects showed the robot's view of the world with the objects marked as "private" blurred out.

Our five measures of interface performance are presented here in our hypotheses. These measures cover interface efficiency (H1), usability (H2, H3), and task performance (H4, H5). A description of how these constructs are operationalized is in Section 4.5.6.

H1 The physical interfaces will **take less time to use for first-time users** than the graphical interface will.

H2 The physical interfaces will **be perceived as more enjoyable and engaging** than the graphical interface will.

H3 The physical interfaces will **be preferred** to the graphical interface.

H4a The physical interfaces will **promote more confidence that privacy settings are honored** than the graphical interface will.

H4b In particular, **the marker interface will promote more confidence than the pointing interface will.**

H5a The physical interfaces will **make it easier to later remember which things are selected** than the graphical interface will.

H5b In particular, **the marker interface will promote better memory than the pointing interface will.**

Most users are familiar with point-and-click graphical interfaces, and clicking seems quicker than physically moving to each object. Then again, a physical interface should take less time to learn, especially for users with poor technical skills.

Our general assumption is that user memory is influenced by (1) whether tags are persistently visible on the objects themselves or only on the monitor, and (2) whether the act of tagging the object is physical or virtual. Physical markers could give users a better mental image of the tagged object, thusly improving both memory and confidence that the right objects were tagged when the user leaves the area. Physically tagging the objects could also aid memory by invoking proprioception. Further, we believe that remembering which objects were tagged will enhance a user's trust and confidence in the system.

Two potential confounds in our system are (1) quality of the interface – i.e., does it appear shoddy or well-designed and (2) responsiveness of the system – i.e., are tagged objects immediately and reliably blurred out and does that blurring persist without flickering?

4.4 Interface Implementation

All three interfaces, illustrated in Figure 4.1, were implemented on a Willow Garage PR2 robot using the Robot Operating System (ROS) [186]. To reduce tracking issues with both the wand and the physical markers the only lighting was the fluorescent lights (the window blinds and door were closed). The robot remained in the same position in the room, with the head (and cameras) always directed at

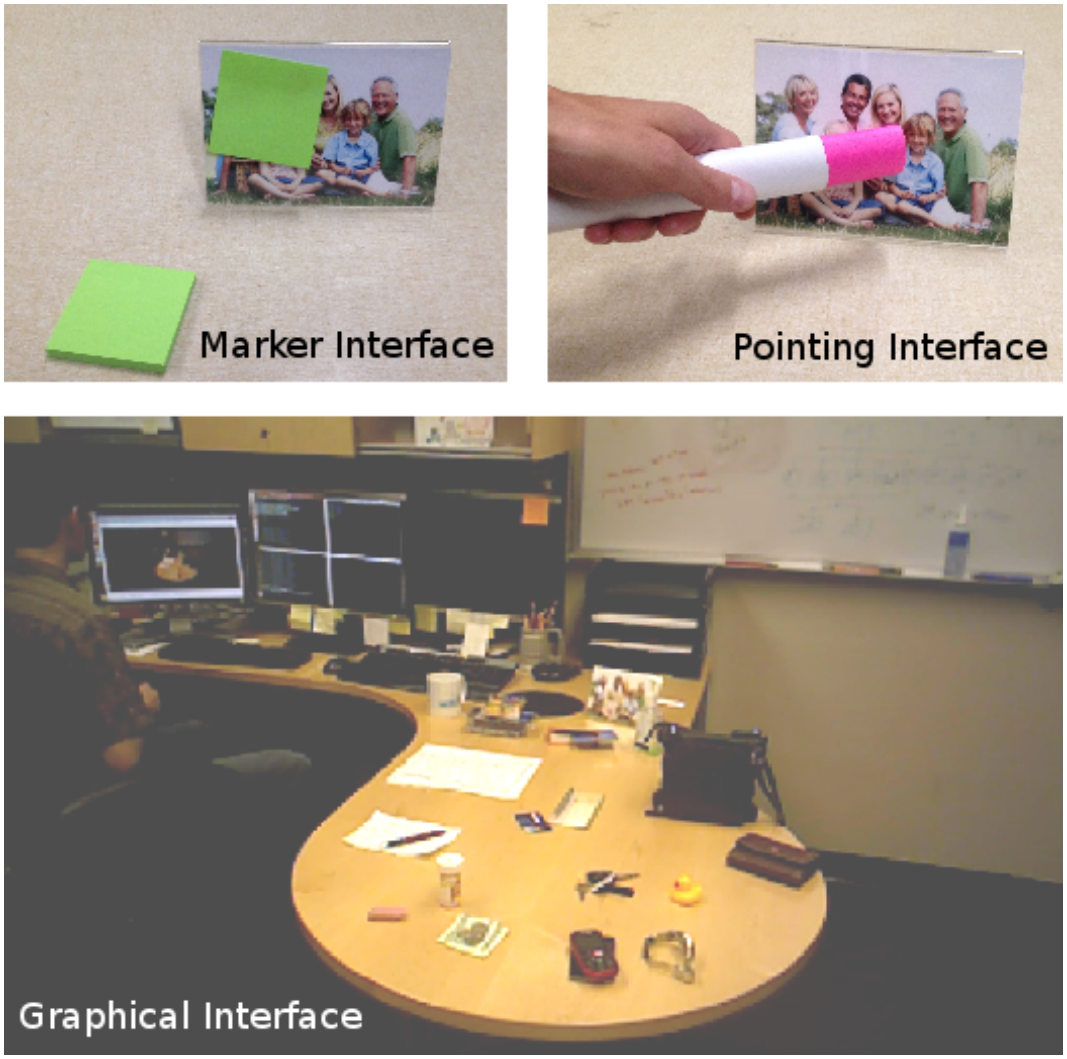

Figure 4.1: The three interfaces used in the study. Top left: marker interface. Top right: pointing interface. Bottom: robot video feed used with a mouse cursor for graphical interface.

the same part of the desk. The automatic gain and white balance on the head-mounted Asus Xtion Pro camera were disabled.

For all three methods object selection consisted of specifying a 3d location ("tag") in the robot's camera frame. A 10cm axis-aligned cube was then centered around that point. All image data inside the projection of that cube was blurred. The corners of the cube were projected onto the screen, with the color indicating the interface method. This approach (over identifying and tracking objects) was chosen to increase reliability. The cube size was selected to cover the largest object without spilling substantially over to other objects. Some participants asked if it was acceptable if the blur partially covered other objects or a big object was not completely covered; this was clarified by the experimenter as being acceptable.

The Marker Interface – The markers used in this study were standard-sized bright green sticky notes. Unlike the wand used in the pointing interface, the markers were identified as tags as soon as they were detected, and tracked persistently.

The Pointing Interface – A so-called "magic wand" tool was given to users for pointing. The wand was a paper cylinder about 27cm long and 3cm in diameter. The final 5.5cm of the length was bright pink, while the rest of the length was white. If the software detected that the wand tip stayed in approximately the same place for two seconds, a tag was added at that location as described above. An object could be untagged by pointing at it again.

The Graphical Interface – In the graphical interface users clicked directly on the robot's video feed in order to tag objects. User clicks were projected into 3d

coordinates in the camera coordinate frame using the PR2's depth camera. An object could be untagged by clicking on the tag again.

4.5 Methods

Each participant tried all three interfaces in turn, first practicing by tagging five objects and then tagging six additional objects designated by the experimenter. This was followed by a freeform task where the participant was asked to tag objects using any combination of interfaces or to physically hide them from the robot's camera to create a "private" office.

The procedure is given in further detail in Section 4.5.7, but first we describe the experimental design and materials.

4.5.1 Study Design

Each participant used and evaluated all three interfaces. This formed a within-subjects design with three conditions, one per interface. To eliminate the confounding effects of order and sequence we counterbalanced the order in which subjects used the interfaces.

4.5.2 Recruitment

Participants were recruited via a flier in a public library and an advertisement on Craigslist. Recruiting on campus was avoided so as not to saturate the sample with

Figure 4.2: Study environment from experimenter's perspective. PR2 robot with gaze fixed on the desk and whiteboard. The 23 target objects were placed throughout the robot's field of view. Participant sat at monitors on the left.

demographics typical among university students. Participants were compensated US$20 for their time (about 1 hour).

4.5.3 Environment

The study was conducted in a single-occupancy office belonging to a faculty member in the College of Engineering. Only the participant, the experimenter, and a PR2 robot were present in the office during the study. Figure 4.2 shows the relative positions of the PR2 robot, computer monitors, and objects. The image was taken from the experimenter's perspective; the participant sat to the left, in front of the monitors. After the study, the participant and experimenter left the office and moved across the hall to a vacant classroom (here, "testing room") to complete the four post-activity surveys. This was to remove the participant from

the office before administering the memory test.

4.5.4 Objects to be Tagged

Twenty-three objects, listed in Table 4.1, were placed on the desk, cabinets, monitors, and whiteboard (always in the same locations). Five objects were selected for timed tagging practice for each interface. These practice objects were chosen for their mundanity and presumed lack of privacy concern. Six additional objects for each interface (18 total) were chosen for tagging after the 5 practice objects. These were chosen from several object classes that we thought would pose privacy concerns (see Table 4.1 for categories). All objects could theoretically be tagged with any of the three interfaces, as well as hidden from view manually.

4.5.5 Other Stimulus Materials

Scenario Description – Participants were told that the robot functions as a janitor, cleaning offices each night after the employees leave. The robot was described as completely autonomous, and its ability to record video was mentioned as a potential privacy concern.

The Cleaning Video – Participants watched a 70-second video of the PR2 robot cleaning the office. Our goal was to provide a realistic mental image of the robotic janitor scenario to engage participants and help them answer our questions about privacy. The robot makes frequent eye contact with the video camera to evoke

Object	**Interface**	**Category**
Mug	All (practice)	–
Rubber Duck	All (practice)	–
Pink Eraser	All (practice)	–
Hand Sanitizer	All (practice)	–
Robot Drawing	All (practice)	–
Pill Bottle	Markers	Embarrassing
Black Purse	Markers	Valuable
Kid's Drawing	Markers	Family/Romantic
Credit Card	Markers	Personal Info
Suggestive Pop-Up	Markers	Embarrassing
Family Photo	Markers	Family/Romantic
Embarrassing Note	Pointing	Embarrassing
Bible and Tract	Pointing	
Junk Food	Pointing	Embarrassing
Cell Phone	Pointing	Valuable
Romantic Note	Pointing	Family/Romantic
Brown Wallet	Pointing	Valuable
Watch	Graphical	Valuable
Cash Money	Graphical	Valuable
Personal Info	Graphical	Personal Info
Tax Forms	Graphical	Personal Info
Checkbook	Graphical	Personal Info
Car Keys	Graphical	Valuable

Table 4.1: Objects used in the study, showing which were used in the practice task and which were the memory objects for the three interfaces. Most objects are also given subjectively-assigned categories, e.g., valuable, financial, personal. Note that the robot drawing and personal info were on the whiteboard, the suggestive pop-up and romantic note were on computer monitors, and the kid's drawing was taped to a cabinet above the monitors.

a social reaction. The video also features playful and old-fashioned background music.

"The Nod" – The robot made approximate eye contact with the participant, gave a slow, sagely head nod, and then fixed its gaze down onto the desk for the rest of the trial.

The Instructional Videos – Each interface condition was preceded by an instructional video of 30–60 seconds that demonstrated to the participant how to use the interface.

Prompt Sheets – Prompt sheets of objects to tag contained a photo and nickname for each object. Each object was photographed in its nominal position on the desk with a rectangle drawn around it for clarity.

Video Feed on Monitor – Throughout the tagging tasks, participants were provided with a live video feed from the Asus Xtion Pro RGB-D camera mounted on the robot's head. As the participant added tags, the tagged objects were blurred on the same monitor (see Section 4.4).

4.5.6 Dependent Measures

Many user studies have relied heavily on self-report measures. Such measures are only accurate if subjects are truly able to introspect on the processes in question. Nisbett and Wilson have famously argued that people only *appear* to do this introspection, and will answer wrongly when their true responses are not what they would have expected *a priori* [171]. To address this potential problem, an effort

was made to validate self-report measures with behavioral measures for several of our hypotheses, as we describe below.

Demographics – We used three surveys to measure individual differences between participants. The first was a custom survey that asked for age, sex, education completed, experience with both robots and household pets, and typical usage of cell phones and social media applications. We also used the Negative Attitudes toward Robots Scale (NARS) developed by Nomura et al. [173] and the three scales about Online Privacy developed by Buchanan et al. [37]. This information was primarily intended to help explain participants' responses to the private objects and to the robot itself, which are not analyzed in this work.

H1: (Efficiency) Time – We timed each participant tagging the same five practice objects with each of the three interfaces. This time included asking questions, reading the list of objects on the prompt sheet, moving to each of the five practice objects, placing a tag, and confirming that the object was blurred on the monitor. For the graphical interface, participants did not need to move between the objects, but were not restricted from looking at the real objects if they were hard to find in the video feed. This yielded three time measurements per participant, one per interface.

H2: (Usability) Enjoyment – At the end of each interface condition participants were asked for feedback via a questionnaire. First, they were given eight items, each with a five-point Likert-type response format anchored at "not at all" and "very much so," as shown in Table 4.2, items 1-8. Example questions are, "Using this interface was tedious," "I felt that the instructions for using this interface

FEEDBACK SURVEY ITEMS FOR EACH INTERFACE

Item No.	**Question (paraphrased)**	**Group**
0	Which interface did you use?	
1	Easy to use	Enjoyment, $\alpha = .83$
3	Fun	
4	Tedious*	
5	A pain to use every week*	
6	A chore*	
2	Instructions could be improved	
7	I used it correctly	Mastery, $\alpha = .60$
8	Could teach to an elderly person	
9	Estimate your practice time	

* Reverse-scored item

Table 4.2: Data reduction results for interface feedback.

could use some improvement," and, "How confident would you feel about teaching an elderly person with poor computer skills to use this interface?" We also asked participants to estimate their elapsed time for the tagging practice task.

H3: (Usability) Preference – After using all three interfaces, participants were asked which interface they would choose to protect their privacy if, "a robot cleans your office every day." This was our primary measure of preference, but did not measure *how much* the participant preferred the chosen interface over each of the other two interfaces. To measure this, a separate question asked the user to assign a dollar value they would be willing to pay to purchase the interface (normalized against the cost of the robot). This measured a participant's *willingness to pay* (WTP) for each interface [6]. The differences between the WTP numbers were intended to measure a user's relative preferences between each pair of interfaces.

H4: (Task) Confidence – Users' confidence that the system would protect their privacy was first measured using the freeform tagging task (described in detail in Section 4.5.7). The user was given the following options for each object: tag it with one of the three interfaces, protect it some other way (e.g., hiding or erasing), or leave it unprotected. This was primarily intended to measure which objects were considered private, but it also indirectly measured the user's confidence that the interface would protect the object's privacy.

This behavioral measure was supplemented by three post-activity survey items. Each item was of the form, "Imagine that you used one of the three interfaces to tag these items: mug, stapler, whiteboard, computer monitor, car keys. How confident would you feel *right now* that those items are reliably blurred out by the robot... using the Pointing interface?" This question was asked three times consecutively, once per interface, in the same order each time. Responses were of a five-point Likert-type format anchored at "not at all" and "completely."

H5: (Task) Memory – We measured each participant's ability to recall the 6 extra objects tagged with each interface, as well as which interface was used for which objects. Participants were given a sheet of paper with four empty boxes labeled, "Physical Markers," "Physical Gestures (the Wand)," "Graphical User Interface (on the screen)," and, "I don't remember which interface I used to tag this object." Participants were given 5 minutes to write down as many objects as they were able to in the boxes. The practice objects were not part of this test, and were listed to the participant out loud by the experimenter so that they would not be confused for the objects in question. Correctly recalling objects and associating

them with the correct interface should tap how memorable the tagging experience was with that interface.

4.5.7 Procedure

The participant was escorted into the office (see Figure 4.2) and seated at the desk. Consent conversation took place with participant sitting in the desk chair. Three pre-activity surveys were administered, which ask about (1) general demographics; (2) attitudes towards robots; and (3) attitudes towards online privacy. The participant was introduced to the janitorial robot scenario via the script and cleaning video. The researcher cued the robot to nod at the participant, which was described beforehand as a "greeting." The robot's head was pointed at the objects on the desk and remained there, stationary, for the remainder of the study.

For each interface, the participant was taught how to use the interface via a short instructional video. In the graphical interface session the researcher also added additional information about the particularities of the RViz implementation. Once this instruction ended, the researcher started the stopwatch for that interface's training session. The tagging prompt sheets guided the participant through tagging the five practice objects. Participants were reminded to check that tagged objects actually became blurred in the robot's video feed on the computer monitor. When the participant reported being done, the researcher recorded the elapsed practice time and administered the second prompt sheet with six additional items to tag. When the participant had tagged all eleven objects, the researcher admin-

istered the interface feedback survey and cleared all the tags before continuing to the next interface.

After all three interfaces had been used, the participant was briefed on the freeform tagging task. The participant was given five minutes to tag or arrange all of the objects mentioned so that he or she felt the desk was acceptably private for an overnight cleaning by the robot. They were told they may use one (or more) of the interfaces and move or change the objects, as long as they didn't damage anything.

After tagging was complete the participant was escorted to the testing room across the hall. Four post-activity measures were administered: (1) the interface confidence survey; (2) the memory test; (3) a survey that asked about the specific objects the participant tagged (not discussed in this paper); and (4) a survey about user impressions of the robot and study scenario (also not analyzed here). After this, the participant was debriefed, thanked, paid, and dismissed. Debriefing cleared up any misconceptions that may have been formed about the robot; namely, we made it clear that the robot could not really clean a room on its own, was not recording anything during the study, and was not being remotely controlled by anyone. The objects were reset to the same nominal locations for the next participant.

4.6 Results

4.6.1 General Demographics

Twenty-seven people (12 female and 15 male) participated in this study. Their ages ranged from 18 to 70, with a mean of 35.2 years. All but three had experience living with pets, and experience with robots varied widely. Level of education, social media usage, and cell phone usage were diverse, indicating different levels of competency and acceptance of newer technologies.

4.6.2 Practice Times (H1)

Since participants used all three interfaces, a repeated measures Analysis of Variance (ANOVA) was used to determine whether they differed significantly in the time spent with each. As mentioned in Section 4.5.1, our conditions were counterbalanced for order, so none of our interface effect checks could be confounded by an order effect. Elapsed times for each interface are summarized in Figure 4.3. There was a significant effect of interface on time (N = 27, $F(2,52) = 34.4$, $\eta^2_{partial} = .570$, $p < .001$). The graphical interface took significantly ($p < .001$) less time to use (M = 36.4s, SD = 21.6s) than both the pointing interface (M = 71.6s, SD = 17.6s) and the marker interface (M = 62.2s, SD = 19.5s). These results were significant even when the two-second delay between placing tags with the pointing interface was removed from those practice times.

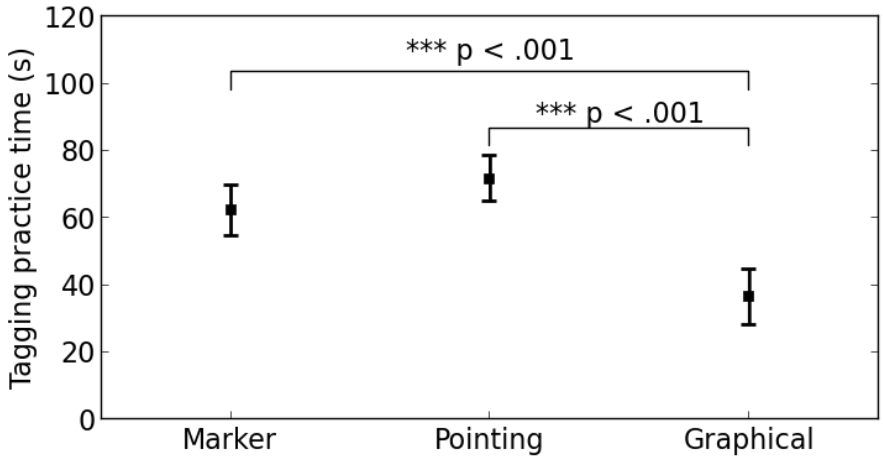

Figure 4.3: Mean practice times with 95% confidence intervals. Significance levels were calculated using a repeated measures ANOVA with the conservative Bonferroni adjustment for multiple comparisons. There was also a significant order effect (not shown in this figure): the first interface took longer than the others.

4.6.3 Enjoyment and Engagement (H2)

Table 4.2 paraphrases each item on our survey. We combined five items to assess "enjoyment" (Cronbach's $\alpha = .83$) and two items for "mastery" ($\alpha = .60$) because they were too highly correlated to report as independent findings. Reports of enjoyment and mastery were not significantly correlated ($r = .16$, $p = .158$). Although the three interfaces did not differ in perceived mastery of them, there were differences in enjoyment (N = 27, $F(2,52) = 6.55$, $\eta^2_{partial} = .201$, $p < .005$). Figure 4.4 displays the mean level of enjoyment across the three interfaces. Overall enjoyment was high but subjects enjoyed using the graphical interface more than each of the others (Marker M = 4.22, SD = 0.732; Pointing M = 4.07, SD = 0.767; Graphical M = 4.55, SD = 0.515).

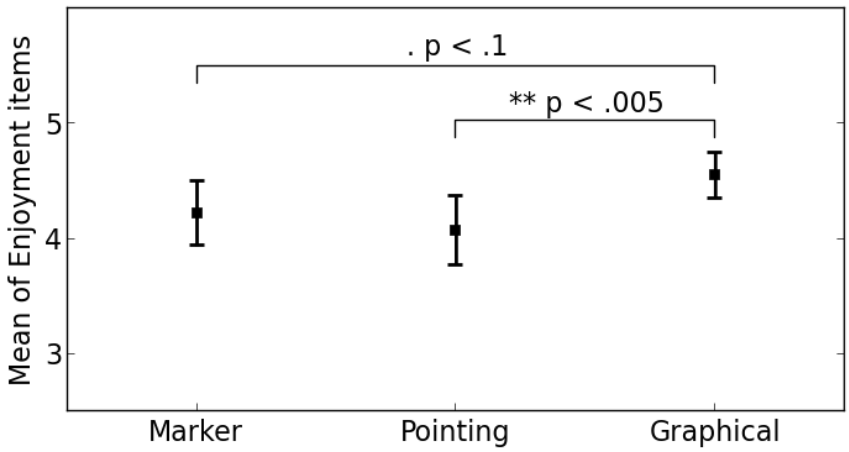

Figure 4.4: Mean enjoyment values with 95% confidence intervals. Significance levels were calculated using a repeated measures ANOVA with the conservative Bonferroni adjustment for multiple comparisons.

The perceived practice times were only slightly correlated with the actual times ($r = .24$, $p = .030$) and showed no significant correlation with enjoyment data ($r = -.07$, $p = .534$). This suggests that the perceived practice times did not measure enjoyment in this study.

4.6.4 Self-Reported Interface Preference: Simple Vote (H3)

Self-reported interface preference was first measured as a simple forced choice between the three interfaces. A chi-square analysis revealed that preferences were significantly different (two-tailed $\chi^2(2) = 6.22$, $p < .05$). In particular, the graphical interface was preferred by over half of our participants (Marker 26%, Pointing 19%, Graphical 55%). None of our demographic variables impacted this preference.

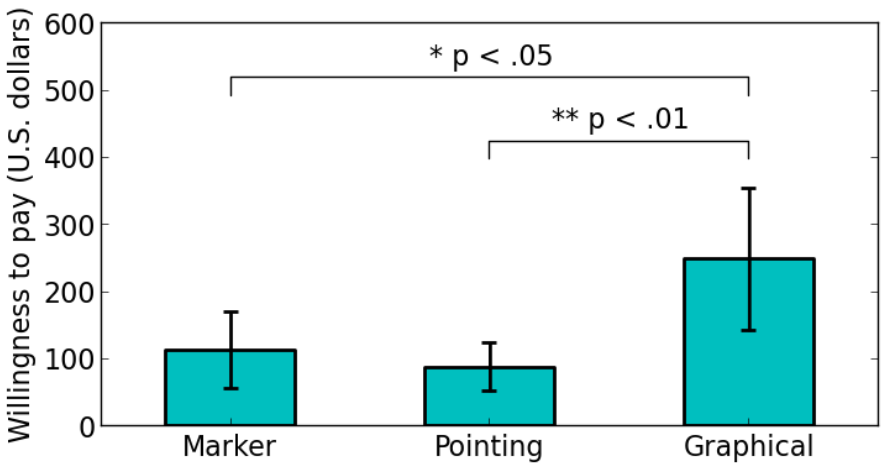

Figure 4.5: Mean willingness to pay (WTP) for each interface with 95% confidence intervals. Significance levels were calculated using a repeated measures ANOVA with the conservative Bonferroni adjustment for multiple comparisons.

4.6.5 Self-Reported Interface Preference: Willingness to Pay (H3)

Results for willingness to pay, our second measure of user preference, are shown in Figure 4.5. The interface effect was significant (N = 27, $F(2,52) = 9.13$, $\eta^2_{partial} = .260$, $p < .001$). Participants were willing to pay more for the graphical interface than for the two physical interfaces (Marker M = \$113, SD = \$147; Pointing M = \$87.8, SD = \$94.7; Graphical M = \$248, SD = \$277). Examining the relative WTP responses for each participant did not reveal a clearly-preferred runner-up to the graphical interface.

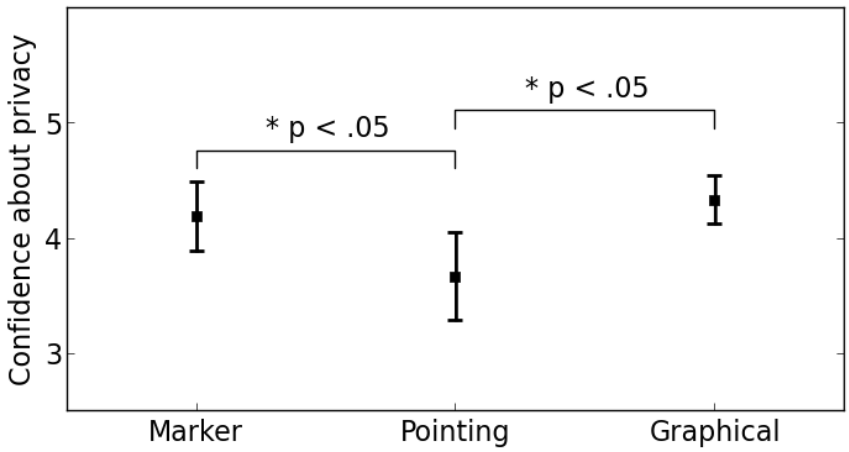

Figure 4.6: Mean interface confidence with 95% confidence intervals. Significance levels were calculated using a repeated measures ANOVA with the conservative Bonferroni adjustment for multiple comparisons.

4.6.6 Self-Reported Confidence in the Interfaces (H4a,b)

Self-reported confidence scores are shown in Figure 4.6. The interface effect was significant (N = 27, $F(2,52) = 6.78$, $\eta^2_{partial} = .207$, $p < .005$). The graphical and marker interfaces were each rated with more confidence than the pointing interface (Marker M = 4.19, SD = 0.786; Pointing M = 3.67, SD = 1.00, Graphical M = 4.33, SD = 0.555). Note that the correlation between confidence and willingness to pay (WTP) was appreciable, although low enough to conclude that they measured different constructs ($r = .328$, $p < .01$).

4.6.7 Freeform Tagging Task (H4a,b)

The freeform tagging task afforded users the opportunity to first choose which objects were private enough to protect in some way, and then to choose an interface or another means for protecting it. Of the 23 objects, an average of 9.2 were protected in some way by a given user, of which 6.6 on average were tagged with one of the three interfaces as opposed to being hidden from view. We computed the fraction of *protected* objects that were tagged with each interface (Marker M = 23.1%, SD = 38.6%; Pointing M = 14.0%, SD = 30.6%; Graphical M = 35.6%, SD = 42.4%), but no significant difference was found between the interfaces.

The pointing interface did not work on the two whiteboard objects, and no users tagged either of those objects with that interface. We checked whether objects were tagged with the interface that was easiest to use on that object (especially whether the markers were avoided for objects that did not offer an easy place to attach the sticky note), but this did not appear to be the case.

4.6.8 Memory of Object Tagging (H5a,b)

For the memory test we counted how many objects were listed in the box for the correct interface. Here we ignored User 12, for whom we accidentally swapped the prompt sheets for the markers and graphical interfaces. The interface effect was significant (N = 26, $F(2,50) = 25.5$, $\eta^2_{partial} = .505$, $p < .001$). Figure 4.7 shows that the six objects tagged with the marker interface were remembered more than those tagged with the pointing and graphical interfaces (N = 26; Marker M = 3.08

objects recalled, SD = 1.47; Pointing M = 1.08, SD = 1.38; Graphical M = 1.38, SD = 1.39).

We were able to use our data to address some possible confounds. In particular, we realized that the 18 non-practice objects ought to have been randomly assigned to the conditions so that object saliency could not confound any effect of interface on recall. We checked whether object saliency could have caused the apparent interface effect, e.g., if the most salient objects happened to be in the marker interface group. Upon inspection of Table 4.1, however, the objects seem to be well-distributed in terms of saliency. Nevertheless, we tried removing the most-remembered object from each group, and the interface effect remained stable.

We also checked the freeform task results to see if it interfered with the memory test, but results from the two tasks appeared to be uncorrelated. Finally, we checked for a sequence effect, e.g., whether tagging with the last interface was easiest to remember. This did not appear to be the case, and would have been dealt with by the counterbalanced order of conditions anyway.

4.7 Discussion

We theorize that practice times were driven largely by the amount of bodily motion required to physically perform the tagging task. The marker and pointing interfaces required whole-arm motions, standing up, and walking to the whiteboard, while the graphical interface required much smaller motions with the computer mouse. This would mean that practice times would vary differently by interface in different

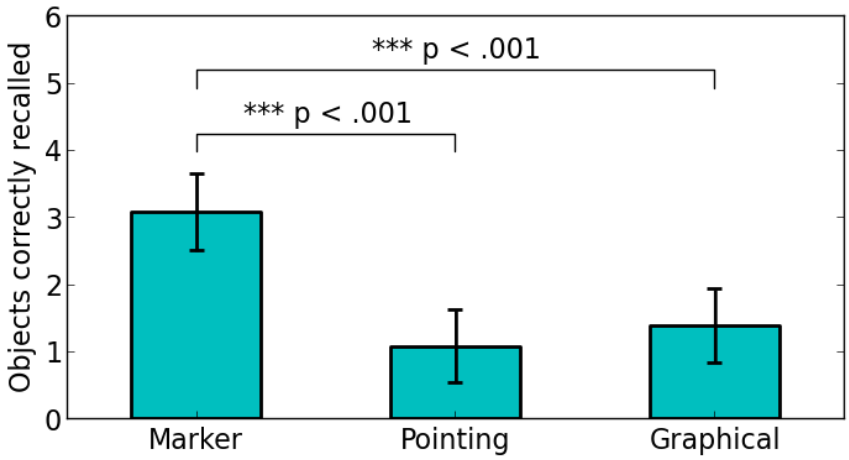

Figure 4.7: Mean true positive rates from the memory test with 95% confidence intervals. Significance levels were calculated using a repeated measures ANOVA with the conservative Bonferroni adjustment for multiple comparisons.

scenarios. For example, the two physical interfaces might require less time if the objects were all within arm's reach. Also, we made the graphical interface easier to use by placing all the objects within the robot's field of view; it might take much longer if the robot had to be teleoperated around the room or even multiple rooms to do the tagging.

The pointing interface consistently failed to tag the two objects on the white-board. Participant feedback from debriefings indicates that user confidence and preference were impacted by this malfunction. We think this caused the interface effect found in the Likert-type interface confidence questions (shown in Figure 4.6). It could have also reduced user preference for the pointing interface shown, e.g., in Figure 4.5. This emphasizes the importance of studying the impact of technical failures on user confidence, which we discuss as future work below.

Some participants noted that the marker interface was unique because the markers could be placed without the robot being present. Those same participants had the insight to worry that, if the robot needs to look at the target objects before they are blurred for the pointing and graphical interfaces to work, then the user's privacy could already be compromised if a remote person is watching through the robot's cameras. Some participants said this impacted their interface preferences, and it may have impacted the preferences and confidence ratings of other participants, too.

4.7.1 Design Implications

Performing this study yielded several best practices for designing privacy specification interfaces:

Feedback – When using the two physical interfaces, people did not like having to look at the monitor to confirm each tag. Add feedback to the tagging tool by placing lights or speakers on the marker or wand, or by having the robot project something onto the scene.

Size and Occlusion – Marker detection should be done in the real, three-dimensional frame instead of image space. This will allow for checking whether a marker candidate is of the appropriate size, regardless of range. Also, 3d sensing and mapping could enable line-of-sight checks so that the system will not erroneously remove a tag that has simply been occluded by a person or other obstacle.

Detection Range – When markers are far away, they are too small to confirm as valid blobs. Since this means we cannot guarantee that no new tag has been placed beyond the effective marker detection range, perhaps the system should blur or redact all pixels beyond a specified range horizon. This would require a depth sensor. In that same vein, larger markers are better as long as they are not easily occluded.

Choice of Marker – Color markers are not robust to lighting conditions, occlusion, or other things of the same color. ARTags are another option, but must be quite large in order to be detected reliably, have a limited viewing angle, and are difficult to distinguish between for humans. We argue that a privacy specification interface would require a much more reliable marker-detector combination in order to be acceptable to users.

Gesture Timing – Participants found that touching each object for two seconds with the wand tool felt tedious. Future interfaces should reliably detect tagging gestures even if those gestures are very brief.

4.7.2 Limitations and Future Work

We received many reports about the negative impact on confidence caused by the limited range of the pointing interface. This suggests that future work should study the impact of technical malfunctions and limitations on user confidence about privacy protection. For example, one could design a study that manipulates the fidelity of the privacy tags—e.g., by flickering the blur effect—coupled

with a Wizard of Oz technique if needed to guarantee the absence of uncontrolled malfunctions.

Feedback from the robot seemed to be critical to whether users trust the system. This feedback could express when tags are added or removed, as well as which filters are in effect. Examining different modes of feedback and the trust (even the *false* trust) they instill in users would be novel and important future work.

The apparent quality of each interface (i.e., whether it appeared shoddy or well-designed) may have impacted user confidence that it actually worked. A future study could manipulate the perceived quality of an interface by varying its appearance and the way it is described by the experimenter. This could reveal an important effect of apparent quality on user confidence.

People often choose between products without having tried them, sometimes using only television commercials or other advertisements to make a choice. Our study only tested our hypotheses on (first-time) interface users, but neglected two other populations: non-users who are given only a description of the interfaces, and long-time interface users. Understanding how these populations might differ and communicate could give nuanced insight into the social acceptance of privacy specification interfaces.

4.8 Conclusions

We hypothesized that physical markers are a better solution for specifying user privacy preferences than GUIs. Our controlled user study has shown that the

reality is more complex.

H1 was reversed – Physical interfaces took more time to use than the graphical interface did.

H2 was reversed – Physical interfaces were perceived as less engaging and fun than the graphical interface was.

H3 was reversed – Physical interfaces were not preferred to the graphical interface; instead, the reverse was true.

H4a was split, but H4b was confirmed – The marker and graphical interfaces promoted more confidence that privacy settings are honored than the pointing interface did.

H5a was split, but H5b was confirmed – The marker interface made it easier to later remember which things were tagged than the pointing and graphical interfaces did. This suggests that *persistence* serves as a significant memory aid.

The graphical interface seems best for applications wherein usability is most important. The marker interface helped users remember what was tagged, which we believe would increase user confidence about privacy. The marker interface also would not require users to grant the robot an unfiltered view of an object in order to tag it as private. The pointing interface had some technical problems that should exclude it from consideration until they are fixed. Overall, choosing the right interface for specifying privacy to a robot seems to depend strongly upon the scenario.

Acknowledgment

We would like to thank Lynn Paul for the use of her office, which offered a cozy and natural setting for the study.

4.9 Lessons Learned

We hypothesized that physical markers are a better solution for specifying user privacy preferences than GUIs. Our controlled user study has shown that the reality is more complex. The graphical interface seems best for applications wherein usability is most important. The marker interface helped users remember what was tagged, which we believe would increase user confidence about privacy. Choosing the right interface for specifying privacy to a robot seems to depend strongly upon the scenario.

4.10 Choosing our Next Contribution

After completing this second contribution about communicating users' privacy preferences to a robot we became interested in how these preferences can change. We treated each user's preferences as static in the studies reported above, but Nissenbaum's theory of "contextual integrity" says that the rules for privacy can change based on the situation [172]. We therefore began to search for factors that might have an especially large influence on privacy preferences. During this search we remembered being especially struck by Darling's demonstration [54]

that giving a personal name and back story to a robot changed the way people thought about it. This phenomenon is called contextual "framing" by scholars [24]. It seemed like changing the "frame" within which a person experiences an interaction might do more than just incrementally change their level of privacy concern—it could suggest a completely different interpretation of the situation. We wondered whether framing might be one of the main factors to consider—perhaps even comparable in importance to the morphology of the robot—for understanding how people perceive interactions with robots.

By bringing contextual frames into our research program we wanted to expand our model of privacy preferences to be more realistic than the simple one we used in this chapter. When we began this (the second) contribution we thought about people as if they have a single, unchanging preference about visual privacy protection for each object or class of objects, like "blur out all tax documents" or "it is OK not to blur that photo." For this reason, our implicit representation of a person's visual privacy preferences was a list of objects with a filtering level next to each object. If our new understanding of the importance of contextual frames was correct, however, each object's privacy sensitivity—one could almost say its *identity* in the eyes of the observer—could depend on the current frame. This would mean that a robot could no longer predict someone's privacy preference for an object just by knowing what category of object it is. The next (third) contribution explores whether this is really the case.

5 Framing Effects on Privacy Concerns about a Home Telepresence Robot

This chapter includes work by Rueben, Bernieri, Grimm, and Smart [199] published in the Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2017).

Abstract– Privacy-sensitive robotics is an emerging area of HRI research. Judgments about privacy would seem to be context-dependent, but none of the promising work on contextual "frames" has focused on privacy concerns. This work studies the impact of contextual "frames" on local users' privacy judgments in a home telepresence setting. Our methodology consists of using an online questionnaire to collect responses to animated videos of a telepresence robot after framing people with an introductory paragraph.

The results of four studies indicate a large effect of manipulating the robot operator's identity between a stranger and a close confidante. It also appears that this framing effect persists throughout several videos. These findings serve to caution HRI researchers that a change in frame could cause their results to fail to replicate or generalize. We also recommend that robots be designed to encourage or discourage certain frames.

5.1 Introduction

Unique to HRI research is the human's interpretation of a scenario; human perceptions and behaviors around robots are unpredictable given only the external, physical facts about the scenario. This work focuses on the impact of frames—i.e., "structure[s] of expectation" [237] within which actions and words will be interpreted differently. It is our intuition that the frame surrounding a given interaction could have comparable or even larger effects on judgments about that interaction than the independent variables typically studied in HRI research, e.g., robot morphology, behavior, environmental factors, and individual differences between subjects. We suspect that even a well-designed HRI can make a bad impression if it is framed such that observers interpret the robot's behaviors negatively; on the other hand, understanding framing effects might be a more efficient way than modifying robot appearance and behavior for reducing or reversing negative reactions to robots. Groom et al. [94] observe that "[HRI] researchers have largely ignored studying framing as an independent variable." We seek to reverse this trend: the studies presented in this paper use framing to manipulate each subject's relationship with the robot operator via short operator biographies in (uniquely) a telepresence scenario.

Our study of framing effects is motivated by privacy concerns in HRI. Privacy is important in all human cultures [11], although different cultures have different norms for privacy and different mechanisms for enforcing those norms. We use the word "privacy" to describe a bundle of constructs related to perceived control

over informational, physical, psychological, and social aspects of one's life [196]. It seems clear that telepresence robots, like autonomous robots, cause concerns about privacy. Telepresence robots are essentially video media spaces, which have a slew of privacy problems themselves (see Boyle et al. [36] for a review), but also mobile, which adds new privacy concerns. Telepresence robots can be driven into private spaces, or used to look around at things against the will of the local user(s). Our broad research focus is on how privacy judgments work in robot-mediated communication; we suspect that framing is among the main factors.

The goal of our research is to answer the question, "how does framing impact privacy judgments?" We have run four human-subjects experiments in rapid succession to gather the first measurements of these effects. Our approach is to use some text to frame a scenario presented via an animated video, to which subjects respond via a questionnaire. We recruit subjects using Amazon Mechanical Turk for quick development and turnaround, beginning with simpler scenarios and variables and then progressing to the privacy concerns that motivate our research.

Each study builds on the previous ones. *Study 1* tests whether we can measure a framing effect with our approach. *Study 2* expands upon the first one to include the effect of re-framing people partway through the study. *Study 3* tests whether our findings generalize to a new video and whether different demographic subgroups are differentially affected by framing. *Study 4* uses a suite of high-fidelity animations specifically designed to evoke privacy concerns and is our first direct measurement of privacy constructs.

This is the first study we know of about framing effects on privacy judgments

in a HRI. Just a few studies have been done in privacy-sensitive robotics; although two [40, 109] focus explicitly on remotely-operated robots, none focuses on the effects of framing or on the way the robot's actions appear from a third-person perspective instead of from its video feed. Past studies of framing in human-robot interactions focus on, e.g., user perceptions of the robot's social role and degree of anthropomorphism, but not privacy, and we are unaware of any framing studies of telepresence robots.

5.2 Related Work

5.2.1 Privacy

Privacy is important in all human cultures [11], although different cultures have different norms for privacy and different mechanisms for enforcing those norms. Prominent theories that describe privacy include Altman's [10] and Nissenbaum's [172]. We use a privacy taxonomy compiled from the literature [196] to divide "privacy" into component ideas: (1) Informational privacy, over personal information, includes (a) Invasion, (b) Collection, (c) Processing, and (d) Dissemination; (2) Physical privacy, over personal space or territory, includes (a) Personal Space, (b) Territoriality, and (c) Modesty; (3) Psychological privacy, over thoughts and values, includes (a) Interrogation and (b) Psychological Distance; (4) Social privacy, over interactions with others and influence from them, includes (a) Association, (b) Crowding/Isolation, (c) Surveillance, (d) Solitude, (e) Intimacy, (f) Anonymity,

and (g) Reserve. We used this taxonomy to define the extents of the idea of privacy so we can work towards covering it all in our research program (i.e., towards *content validity*). The small amount of work so far on "privacy-sensitive robotics" includes Hubers et al. [109], Butler et al. [40], and Rueben et al. [198].

5.2.2 Privacy Concerns about Telepresence Systems

Robot-mediated communication has become possible for doctors, workers, bosses, and visitors to older adults [157]. These telepresence robots create interactions that differ both from face-to-face interactions [36] and from purely virtual systems like avatar-based telepresence [28, 29, 141]. Even video media spaces change the privacy situation because the remote operator's actions are seen outside his/her context and dissociated from his/her identity [36]. Giving the remote user a physical (robot) body raises additional normative questions, like whether it is acceptable to rest one's feet on the robot's base [141]. Several studies on telepresence robots for older adults have identified privacy concerns [26] and behavior changes due to feeling watched by the robot [43]. The paradigm shift expected from the advent of telepresence robots may even prompt changes to U.S. privacy law [119]. This appears the be the first study that focuses on both privacy concerns and telepresence robots (in Section 5.7).

5.2.3 Animations for Studying HRI

Several studies have compared human-robot interactions over video to live ones. Woods et al. [258] have shown a strong agreement in general between being approached by a real robot and by a robot in a video. The same authors also cite findings by developmental psychologists, however, that "while babies happily interact with their mothers via live video, they get highly distressed when watching pre-recorded or replayed videos of their mothers (as it lacks the contingency between mother's and baby's behaviour)" [257]. We were careful not to use our videos to study scenarios that normally require interaction.

Powers et al. [184] compared interactions with an animated computer agent against interactions with a real robot over a video feed. They found that engagement was higher and positive personality traits were more strongly associated with the real robot, but also that people remembered what the agent said better than what the robot said. Also, McDonnell et al. [160] compared emotive actions performed by a human actor to the same actions mapped onto animated bodies. They found that the perception of emotions in the actions was mostly the same across body conditions.

Our methodology resembles the one used by Takayama et al. [236], which prototypes robot behaviors as animations and shows the videos to a large sample of people online. We use Amazon Mechanical Turk (MTurk) to recruit our subjects—consult Mason and Suri [159] for some studies that compare MTurk users to laboratory subjects. Performing a study with animated robot behaviors allowed us to

provide consistent experiences for each of the study participants, to test a variety of task domains, and to engage a geographically diverse set of study participants. In terms of design research, using animations allowed us to test the behaviors we would like to build before locking in the design a robot would need to physically perform the behaviors.

5.2.4 Framing

A frame is a "structure of expectation" [237] within which actions and words will be interpreted differently. Framing language is metacommunicative; it tells one the frame in which to interpret subsequent communications [24]. Bateson [24] gives the example of monkeys engaged in a playful fight; the monkeys know that the bite that would normally be aggressive is fun in this context. Here the frame is "play", and "this is 'play'" would be framing language. Tannen [237] reviews the idea of a "frame" as well as the related terms "schema" and "script" across disciplines.

Groom et al. [94] cite some studies of existing expectations about robots, but observe that "researchers have largely ignored studying framing as an independent variable." Their study manipulates the role of a robot, an instance of framing, in a search and rescue context. Howley et al. [106] also manipulate the robot's social role, whereas Fischer et al. [72] have a robot issue a greeting to frame the interaction as social. Paepcke and Takayama [177] manipulate user expectations about the robot's capabilities, a framing that is very relevant to the concerns of Richards and Smart [191]. Darling [54] uses a narrative about the robot to

manipulate how much users anthropomorphize it. We consider all these studies to use framing. The studies presented in this paper use framing to manipulate the familiarity of the robot operator to the participant via short operator biographies in (uniquely) a telepresence scenario.

5.3 Approach

We use the same five-part methodology for each of the four studies reported here.

(1) We *frame* our participants by having them read a paragraph of text. This text describes the occasion of the interaction and introduces the robot operator, whose identity we manipulate in our experiments. Our goal was to provide enough context for the video to make sense but leaving the robot's actions ambiguous enough to require interpretation.

(2) Next, we present our main *stimulus*, which is an animated video of a PR2 robot performing some actions inside a home (see Figure 5.1). We chose the PR2 because it has a mobile base and two arms for performing human-like manipulation tasks as well as a head and obvious eyes so users can follow its gaze, but lacks a screen for showing imagery of the remote operator—something that would introduce many variables we don't want to deal with yet. The animated videos were made in Blender. We chose to use animation so we could specify the layout and appearance of the home, and also so that the robot could perform natural-looking actions that are not only beyond the state-of-the-art for autonomy but also difficult to do using teleoperation.

Figure 5.1: Representative thumbnails from the animated videos used for the studies presented in this paper. See Sections 5.4–5.7 for video descriptions.

(3) *Responses* are gathered using a questionnaire. We expect subjects to interpret the videos from within the contextual frame we've given them and respond accordingly.

(4) We *interpret* what is being measured by the survey responses using principal component analysis (PCA).

(5) We *iterate* on this process by using the results from one study to re-design our survey for the next study. Successive studies can act as (at least partial) replications that contribute to a meta-analysis. Iterating is made easy by the response speed on Amazon Mechanical Turk; our studies typically complete within a few hours of being launched. Also, all the elements of our studies are easy to change quickly: a paragraph for framing, an animated video for the main stimulus, and a questionnaire for the response.

5.4 Study 1: Opening the Fridge

Our first study tested whether the framing effect is measurable, and if so how large it is (RQ1-A). In all four studies our two frames manipulate the subject's familiarity with the robot operator within the hypothetical scenario.

5.4.1 Methods

Frame. This study used a between-groups design with two framing conditions. The manipulated variable was the familiarity of the respondent's relationship with

the robot operator in the hypothetical scenario. The operator was "your sister" in one framing and "a home appraiser" whom "you have spoken with... once over the phone, but have never met... before in person" in the other. In both cases we used the name Lisa. A general motivation was provided for each character so the scenario was not confusing: the sister is seeing your remodeled kitchen for the first time, whereas the home appraiser is checking that it's up to code.

Main Stimulus. The video (16s long) shows an animated PR2 robot in the kitchen. The robot opens the refrigerator, looks inside, and closes it again. The camera perspective is from the next room over, the living room.

Survey Items. We created seven items about trust, comfort, and acceptability to measure respondents' concerns about each scenario. They all used a 7-pt Likert-type response format. Two open-ended questions checked for attention and understanding of the two scenarios. No demographic information was collected in the survey.

We recruited 64 people total—31 saw the "sister" condition and 33 the "stranger" (i.e., home appraiser) condition. Subjects in each of the four studies were paid approximately $10 per hour ($1.50 per HIT for this study based on a predicted duration of about 9 minutes) and we always ensured that nobody took the same survey twice.

5.4.2 Results

Data reduction via principal component analysis (PCA) yielded 2 dependent variables: 6 items were combined whereas 1 remained separate, namely, "I think it's important for me to be at home while this is happening." The 6-item composite we named Comfort ($\alpha = .95$); the other item we will call ShouldBeHome for short. Correlation between these two dependent variables was $r(62) = -.37$.

Respondents were more comfortable and felt less need to stay home in the "sister" condition (RQ1-A). Two-sample t-tests showed these framing effects to be statistically significant with medium-high effect sizes (on Comfort: $t(62) = 3.39$, $d = 0.847$, $p = 0.00123$; on ShouldBeHome: $t(62) = -3.16$, $d = -0.790$, $p = .00245$).

Many people described the home appraiser's behavior as "nosy" in the open-ended responses. We decided to use this word in future versions of the survey because it fits the sorts of concerns we tried to evoke in our scenario; apparently many respondents also found it apt.

5.4.3 Discussion

These initial results show that the framing effect is present, sizable, and able to be detected via an online survey with text prompts and an animated video. The next 2 studies measure additional variables while replicating and generalizing the framing effect.

5.5 Study 2: Fridge (Within Subjects)

The second study aimed to test what happens when a person switches between interpretive frames. We did this by re-framing our subjects in the middle of the experiment, resulting in a repeated measures design. First, we sought to replicate the framing effect we detected in Study 1 (RQ2-A). Second, we wondered whether showing both frames in series would reveal any carry-over effects, such as a halo effect that would make judgments more similar or a contrast effect that would make them more different (RQ2-B).

5.5.1 Methods

In this study, each subject saw both framing conditions. The framings and video were the same as for Study 1 except that the home appraiser is now a home "inspector" named "Alice" (your sister is still named "Lisa"). The order of conditions was counterbalanced. In between the two conditions we showed a text note that told respondents that an entirely new scenario will now be presented, but with the same video.

Survey Items. We used the same questions as in Study 1 with some minor wording changes as well as the addition of two new questions about whether Lisa was being "nosy" or "rude," inspired by some open-ended responses in Study 1. We also moved the open-ended items from the end of the questionnaire to its beginning to encourage respondents to think about what they saw before responding to the other items. We added a more explicit manipulation check after both conditions—

"In one sentence, describe the difference between the two scenarios"—to make sure respondents read and understood the framings.

We analyzed responses from 41 people after omitting 5 for failing the manipulation checks—of the 41, 22 saw "sister" first and 19 saw "stranger" first. Eight had also participated in Study 1.

5.5.2 Results

Unlike for Study 1, data reduction here yielded a single dependent variable ($\alpha = .95$). We named this single, 9-item dependent variable Comfort, but it consisted of different items than in Study 1. This data reduction remained stable even when responses were split by condition or order except for one interesting exception: ShouldBeHome was more strongly correlated with BeingRude and BeingNosy in the "stranger" condition than in the "sister" condition.

Respondents were more comfortable with the "sister" condition (RQ2-A). According to an ANOVA, this within-subjects framing effect on Comfort was statistically significant with a large effect size ($F(1,40) = 18.4$, $\eta_p^2 = .315$, $p < .001$). We also checked for an order effect on Comfort, but it was not statistically significant and was much smaller in magnitude ($F(1,40) = 1.34$, $\eta_p^2 = .033$, *n.s.*).

Regarding the carry-over effect, subjects' Comfort ratings were more sensitive to framing when the "sister" condition was viewed first (RQ2-B). A two-sample t-test on the differences in means between the two framing conditions, however,

revealed that this effect was not statistically significant (mean difference in Comfort for "sister" first: 1.73, for "stranger" first: 0.98, $t(39) = 1.16$, $p = .253$).

5.5.3 Discussion

This study successfully replicated the large framing effect from Study 1. The effect size may be different depending on which condition comes first, but more evidence is needed to confirm this. We continue counterbalancing the order of conditions in our subsequent studies to prevent order from confounding our framing manipulation, as well as to continue measuring any difference in effect size based on which frame is presented first.

5.6 Study 3: Playing Chess

We next test whether the framing effect generalizes to a different video and scenario description (RQ3-A). We also conduct a second test of whether the framing effect is moderated by which frame comes first (RQ3-B) and a first test of whether the name of the operator matters (RQ3-C). Finally, we report on the range of subjects' sensitivity to the change in frames (RQ3-D) and test whether some basic demographics are correlated with our dependent variables (RQ3-E),

5.6.1 Methods

Main Stimulus. The video we used for this study shows a new scenario. The PR2 is now with the observer in the living room; it drives around a chess board to face the observer, looks down at a chess board, and makes a move.

Frame. The framing paragraphs that precede the video were rewritten to fit the new scenario, but still manipulated subjects' hypothetical familiarity with the operator: the operator is either a friend you play chess with frequently or a stranger you've only met on a chess website. Also, we counterbalanced the assignment of the operator's name (Lisa or Alice) to each framing condition (sister and stranger).

Survey Items. After each showing of the video, 17 items with Likert-type response formats were presented (paraphrased in Figure 5.2). Of these, 7 were adapted from the 9 items in Study 2; 10 were new, many of which were tailored to this scenario (e.g., "I think using a robot like this is a good idea for improving the online chess experience."). Here we began using a 9-pt response format instead of a 7-pt one to combat floor and ceiling effects.

Next we asked 20 demographics questions. These included general information such as age, sex, and education level as well as items about experience with robots and living situation (e.g., "I am used to sharing my living space with other people such as friends, family, roommates, and guests.").

The 14 NARS items by Nomura et al. [173] were presented next, but they were adapted to be about telepresence robots (e.g., "If robots had emotions, I would be able to make friends with them" became "If robots could express emotions

for someone far away, I feel I could become friends with that person") and we clarified some wordings to address the comments by Syrdal et al. [233]. We will refer to these items with "NATS", which stands for "Negative Attitudes about Telepresence (robots) Scale."

We analyzed 61 out of the 66 responses after excluding 5 people for failing our manipulation checks.

5.6.2 Results

This sample comprised 38 men, 22 women, and 1 person who left that question blank. Most respondents (72%) were white. Mean age was 32.5 years (SD: 9.2 years). Everyone had completed high school and only 4 were unemployed or unable to work. 89% had at least seen a video of a robot before, and 43% had driven a remote controlled vehicle. 75% reported that they use social media daily and 85% that they use a smartphone. Only 36% live alone and only 16% live with children.

We chose to reduce our 17-item video response survey into 3 dependent variables named PositivePresence (7 items, $\alpha = .92$), WasNotTooFast (5 items, $\alpha = .86$), and PilotEtiquette (9 items, $\alpha = .94$). Figure 5.2 shows the items included in each variable and their factor loadings. Correlations between these variables range from $r(120) = +.69$ to $+.84$.

All three variables were higher in the "friend" condition than in the "stranger" condition (RQ3-A). An ANOVA revealed that these framing effects were large and statistically significant (PositivePresence: $F(1,57) = 88.2$, $\eta_p^2 = .607$, $p < .001$;

	PositivePresence	WasNotTooFast	PilotEtiquette
ActedAppropriately	-0.25	0.16	**0.89**
BeingRude	0.13	-0.23	**-0.85**
GoodManners	-0.41	0.07	**0.78**
CameTooClose	0.27	-0.46	**-0.65**
GoodIntentions	**-0.55**	0.28	**0.58**
WorriedSneaky	0.33	**-0.53**	**-0.57**
AfraidForSafety	0.20	**-0.54**	**-0.55**
ComfyDistance	-0.50	0.35	**0.51**
FeelGood	**-0.80**	0.23	0.42
GoodForChess	**-0.79**	0.24	0.23
FeltSheWasPresent	**-0.79**	-0.08	0.04
FineWithLogin	**-0.72**	0.32	0.42
TrustToOperate	**-0.62**	0.39	**0.52**
OKtoLoginLater	**0.53**	-0.38	-0.28
ArmsSlowerPlease	0.16	**-0.83**	-0.15
DriveSlowerPlease	0.12	**-0.82**	-0.24
SelfConscious	0.41	**-0.63**	-0.21

Figure 5.2: Table of factor loadings for data reduction of 17-item video response survey in Study 3. From left to right, we named these PositivePresence, WasNotTooFast, and PilotEtiquette. Items placed in each composite are underlined.

WasNotTooFast: $F(1,57) = 79.9$, $\eta_p^2 = .584$, $p < .001$; and PilotEtiquette: $F(1,57) = 86.8$, $\eta_p^2 = .604$, $p < .001$).

As in Study 2, the framing effect was larger when the "friend" (previously "sister") framing came before the "stranger" framing (RQ3-B). Yet again, however, this effect was not statistically significant (it was largest for WasNotTooFast: difference in means for "friend" first = 1.58, for "stranger" first = 1.07, $t(59) = 1.47$, $p = .146$).

Statistically significant name effects were found for PilotEtiquette ($F(1,57) = 5.82$, $\eta_p^2 = .093$, $p = .019$) and almost for WasNotTooFast ($F(1,57) = 3.54$, $\eta_p^2 = .058$, $p = .065$), but these turned out to be due to a breakdown of random assignment[1] (RQ3-C). None of the three order effects was statistically significant (all ps $> .1$) and observed effect sizes were relatively small (all η_p^2s $< .04$).

There were notable individual differences in sensitivity to the framing manipulation (RQ3-D). For example, PositivePresence ratings changed by a mean of 1.53 points between framing conditions, but a few respondents changed not at all or in the opposite direction. Upon inspection, it looks like many of these people did not demonstrate their understanding of the conditions very well in the open-ended questions. Manipulation checks like these are crucial for gauging ex-

[1]Subjects assigned to the group with Lisa as the friend responded more affirmatively to the item, "I really dislike it when people enter my bedroom without my permission." This imbalance mattered because subjects who responded more affirmatively to that item were also more sensitive to the framing manipulation. Thus, part of the framing effect appeared to be a naming effect. Adding that demographic variable to the ANOVA caused all naming effects to lose statistical significance.

perimental validity in online surveys. Most people were affected by the framing in the expected way, however; the 95% confidence interval for the mean sensitivity of PositivePresence only spans from 1.20 to 1.86 points.

We kept 13 of the 14 items in the NATS and call it the NATS-13 (α = .91). The last item in the scale, which reads, "I feel that in the future society will be dominated by robots", only correlated at $r(59)$ = .19 with the rest of the scale. Since it seems to measure a belief that could be relevant to judgments about our scenario, however, we chose to keep it as a single-item variable called SocietyWillBeDominated.

The NATS-13 was especially strongly correlated with our three dependent variables (PositivePresence $r(59)$ = -.71, r^2 = .50; WasNotTooFast $r(59)$ = -.73, r^2 = .53; PilotEtiquette $r(59)$ = -.71, r^2 = .50), accounting for half of their variance (RQ3-E). Actually, this was a stronger effect than the framing manipulation itself; moving up one point on the NATS-13 was about equivalent to switching the frame from "friend" to "stranger." These correlations were statistically significant even after a conservative Bonferroni correction to account for the heightened risk of Type I error from checking the correlations between 17 of our demographic variables and each of the 3 dependent variables.

Many of our other demographic variables besides the NATS-13 were also correlated with ratings of PositivePresence, WasNotTooFast, and PilotEtiquette (e.g., "In general, I enjoy having guests over to where I live") (RQ3-E). A few were correlated with sensitivity to the framing manipulation (e.g., "I really dislike it when people enter my bedroom without my permission"). It is interesting that these

two things don't always occur together (e.g., the "enjoy having guests" example above was not strongly correlated with sensitivity to the framing manipulation).

5.6.3 Discussion

Our results support the conclusion that the framing effect generalizes to the chess scenario. This study has also yielded much more information than the first two did, including demographic profiles of participants, correlations between demographics and our dependent variables, and first looks at sensitivity and effects of operator name. We now turn to our constructs of interest in the realm of privacy.

5.7 Study 4: Tour before a Party

This study uses four videos made to evoke different privacy concerns on our taxonomy [196] as well as survey items to measure these privacy concerns. Our main research question is whether the framing effect replicates in this domain (RQ4-A). We are also interested in several types of order effects. First, do respondents get used to seeing a robot poking around their house after watching the four videos, causing privacy concerns to be lower in the next condition (RQ4-B)? Second, is the framing effect moderated by which frame comes first (RQ4-C)? Third, does a frame wear off as respondents watch the videos (RQ4-D)? After all, the videos do not show the operator's face or any other explicit cues that he is telepresent.

We will also take another look at whether the demographics we have targeted

are correlated with any of our privacy variables (RQ4-E), as well as at how sensitive these new variables are to the framing manipulation (RQ4-F). Also, we want to know how much of the framing effect is attributable to changes in trust of the operator (RQ4-G). Finally, we look at which of the four videos is most (or least) concerning with respect to privacy (RQ4-H).

5.7.1 Methods

Frame. The scenario is that you have invited some people to a party in your home, but one person can't attend except by logging into the robot. That person is either a close friend whom you see weekly or a stranger whom you met for the first time today. We used male names this time: Will and Chris.

Main Stimulus. The main stimulus for this study was divided into four new animated videos with survey items presented after each video. This way, participants could answer questions directly after watching a certain part of the scenario. The videos were more photorealistic than those used in the previous three studies (see Figure 5.1). They range from 21–34s long. Each is designed to evoke certain privacy constructs in order to cover as much of the taxonomy (see Section 2.1.3 [196]) as possible: e.g., Surveillance and Psychological Distance in the "gaze" video (the robot makes eye contact with you), Invasion and Territoriality in the "desk" video (looks in your desk drawer), Anonymity in the "photo" video (picks up a family photo), and Modesty in the "bedroom" video (sees your messy bedroom, including some women's underwear).

Survey Items. We used all new items for the video response survey. These were designed to target the privacy constructs evoked by the videos, as well as ask some more general questions. All 22 items are paraphrased in Figure 5.3. They all use a 9-pt Likert-type response format. Note that these items, as in the other studies, are context-sensitive; they might tap different constructs if we used them in a different scenario.

Immediately after reading each framing paragraph, participants took a 7-item trust questionnaire. Six items were adapted from the Specific Interpersonal Trust Scales (SITS-M and SITS-F) [115] and one item, shown last, we created: "I trust [Will/Chris]."

The demographics questions were the same as in Study 3 except that the NATS was modified slightly and reduced to 8 items for brevity.

At the very end of the survey we asked our respondents to rank our four videos by how concerning they were with respect to privacy.

This study was counterbalanced for the order of the two framings, for the order of the four videos according to a Latin Square design, and for which name was assigned to which framing as in Study 3. We analyzed 65 responses after discarding one that failed the manipulation check.

5.7.2 Results

Of the 65 responses we analyzed, 40 were males and 25 were females. 62% marked "white" as their ethnicity. The mean age was 32.0 years (SD: 8.8 years). All our

respondents indicated that they had graduated high school, and all but 18% are (self-)employed. 49% had at least driven a remote-controlled vehicle before. 63% use social media every day, and all but 1 owned a smartphone. Only 29% live alone; another 29% live with children at least sometimes.

Data reduction was performed on all 22 video response questions at once to look at how the different privacy constructs might relate. Using a PCA we chose 4 composite variables for reporting (Figure 5.3). Each composite variable combines items that were meant to tap different privacy constructs, indicating either that those constructs overlapped in our respondents' minds or that our items failed to discriminate between them. WorriedAboutLikenesses (7 items; $\alpha = .89$) includes items about dissemination of personal information, anonymity, and psychological distance; DontMessWithMyStuff (5 items; $\alpha = .87$) includes items about invasion of personal information and territory; EmbarrassedByMess (5 items; $\alpha = .90$) includes items about collection and processing of information, namely a messy room that could cause someone to judge you; HardToBeAlone (3 items; $\alpha = .91$) includes items about solitude, intimacy, and surveillance. Intercorrelations between composites ranged between $r(128) = +.60$ and $+.73$.

Two items, one designed to tap surveillance and one for modesty, didn't correlate much with any of the others (all $|r(128)| < .22$). After analysis, we believe that they were not good measures of any relevant constructs to this study, so we do not report on them further here[2].

[2]Neither item had any statistically significant framing effects, which were very large for the rest of our dependent variables, or any statistically significant correlations with our demographic variables. The item texts were "I would feel comfortable doing some small tasks to prepare for

	WorriedAboutLikenesses	DontMessWithMyStuff	EmbarrassedByMess	HardToBeAlone
GAZE_put_it_online	**0.66**	0.15	0.37	0.35
GAZE_eye_contact_discomfort	**0.66**	0.18	0.18	0.35
GAZE_ok_getting_ready	(omitted from PCA)			
GAZE_hard_to_be_alone	0.25	0.23	0.21	**0.85**
GAZE_hard_to_talk_privately	0.28	0.32	0.18	**0.82**
GAZE_filter_myself	0.48	0.20	0.38	**0.60**
DESK_no_drawer_sans_permission	0.05	**0.86**	0.24	0.07
DESK_worried_saw_info	0.43	0.52	**0.57**	0.18
DESK_ok_if_sees_screen	**-0.55**	-0.38	-0.28	-0.13
DESK_not_ok_because_mine	0.14	**0.89**	0.18	0.19
DESK_filter_things	0.38	**0.63**	0.39	0.29
PHOTO_ok_to_see_photo	**-0.76**	-0.09	-0.01	-0.02
PHOTO_worried_will_drop	0.28	**0.59**	-0.05	0.22
PHOTO_concerned_can_recognize	**0.76**	0.03	0.22	0.13
PHOTO_filter_things	**0.70**	0.30	0.02	0.32
BEDROOM_embarrassed_saw_mess	0.23	0.21	**0.86**	0.09
BEDROOM_would_have_cleaned	0.05	0.12	**0.85**	0.25
BEDROOM_worry_will_judge	0.46	0.24	**0.66**	0.15
BEDROOM_worried_will_put_online	**0.72**	0.20	0.34	0.28
BEDROOM_tshirts_over_underwear	(omitted from PCA)			
BEDROOM_bad_to_look_unpermitted	0.09	**0.59**	0.46	0.35
BEDROOM_filter_things	0.32	0.46	**0.51**	0.47

Figure 5.3: Table of factor loadings for data reduction of 22-item video response survey in Study 4. From left to right, we named these WorriedAboutLikenesses, DontMessWithMyStuff, EmbarrassedByMess, and HardToBeAlone. Items placed in each composite are underlined.

The framing manipulation effect was always in the predicted direction for our four privacy variables: higher privacy concerns for a stranger than for a friend (RQ4-A). An ANOVA revealed that these effects were all large and statistically significant (WorriedAboutLikenesses: $F(1,49) = 91.8$, $\eta_p^2 = .652$, $p < .001$; DontMessWithMyStuff: $F(1,49) = 45.5$, $\eta_p^2 = .481$, $p < .001$; EmbarrassedByMess: $F(1,49) = 79.4$, $\eta_p^2 = .618$, $p < .001$; HardToBeAlone: $F(1,49) = 38.9$, $\eta_p^2 = .443$, $p < .001$).

The effect of acclimatization (i.e., first condition vs. second condition) was only statistically significant for DontMessWithMyStuff ($F(1,49) = 4.71$, $\eta_p^2 = .088$, $p = .035$) and almost for HardToBeAlone ($F(1,49) = 3.04$, $\eta_p^2 = .058$, $p = .088$). These were both in the expected direction as well: concerns about the operator messing with your stuff or making it hard for you to be alone were lower in the second condition, suggesting that respondents became complacent as they got used to the scenarios (RQ4-B). The name (Will or Chris) of the operator was also included in the model, but was not found to have any statistically significant effects.

Just like in Studies 2 and 3 we checked whether sensitivity to framing was moderated by which frame came first. There was a consistent effect: the relationship effect was larger on all 4 composites when the respondent saw the "stranger" frame first (RQ4-C). This is in the opposite direction, however, as it was in Studies 2 and 3. To test for statistical significance, we ran two-sample t-tests for each of

---

the party knowing that Will could drive by and look at me like this" ("gaze" video) and "I wouldn't be as upset if the clothes basket contained t-shirts instead of underwear" ("bedroom" video).

the 4 composites between two groups: those who saw "friend" first (n = 31) and those who saw "stranger" first (n = 34). This difference was only statistically significant for DontMessWithMyStuff (mean sensitivity for "friend" first = -0.77, for "stranger" first = -1.56, $t(63) = 2.41$, $d = 0.601$, $p = .019$).

In this study we tested whether the effects of a frame wear off as the respondents watch the four videos. For each video (e.g., "desk") within each frame (e.g., "friend") we ran a MANOVA (8 total) that tested whether that video's position in the order of videos (e.g., 3rd of 4) predicts the responses to the items that go with that video. None of these 8 tests was statistically significant; it appears that our two frames remained active for the extent of each condition, and perhaps could have for much longer, or until a significant distraction occurred (RQ4-D).

We chose 5 out of the 8 NATS items (the "NATS-5"; $\alpha = .81$) for a measure of negative emotions about telepresence systems. The remaining three we kept as single-item variables. The first was "If robots could express emotions for someone far away, I feel I could become friends with that person" (correlations with the other 3 variables were all $|r(63)| < .43$). The other two were written to figure out why the item that said "I feel that in the future society will be dominated by robots" didn't fit into the rest of the scale in Study 3; one reads, "I feel that in the future robots will be everywhere in our society", and the other reads, "I feel that in the future society will be controlled by robots." The "society will be controlled" item appears to be the misfit: it was not correlated with the NATS-5 ($r(63) = -.04$). The "robots will be everywhere" item was negatively correlated ($r(63) = -.46$) with the NATS-5, so it appears that discomfort with telepresence

was linked in our sample with a belief that robots will not become ubiquitous anytime soon.

Some of our demographic variables were correlated with our privacy DVs (RQ4-E). The largest correlation was $r(128) = +.40$ between a composite of CleanUpBeforeGuests and CloseDoorsBeforeGuests ($\alpha = .49$) and EmbarrassedByMess. Also, the NATS-5 was correlated with WorriedAboutLikenesses at $r(128) = +.36$ and also the other 3 main composites, but those correlations—and, in fact, all other IV-DV correlations—were not statistically significant when we protect for Type I error from checking all possible correlations. Note that the NATS items account for much less variance here than in Study 3, wherein the 13-item NATS composite accounted for half of the variance of the dependent variables.

We also looked at how sensitive our respondents' privacy concerns were to the relationship manipulation (RQ4-F). In this study, the variable with the most individual differences in sensitivity was HardToBeAlone—its 95% confidence interval spanned from a sensitivity of -1.99 to -1.02. None of the confidence intervals for our 4 composite privacy variables crossed zero. There were also some sizable correlations between demographic variables and these sensitivities, especially with the living situation and NATS variables, but none reached statistical significance after protecting for Type I error.

According to a PCA, the 7-item scale about trusting the operator measured a single dimension, which we called "Trust" ($\alpha = .95$). It is interesting to note that although the SITS [115] from which we drew all but 1 of the items claims to measure multiple dimensions of trust, with slightly different dimensions between

men and women, our items only appear to have measured one construct for both sexes in this study. We propose that this is because the respondents do not know much about the robot operator in either framing condition, so their judgments about him are not as complex as they might be for, e.g., a close friend in real life.

We can validate our composite measure of Trust by testing whether it is manipulated by the difference in frames ("friend" vs. "stranger") and not by the difference in names ("Will" vs. "Chris"). Matched-pair t-tests supported the validity of our Trust measure (relationship effect: $t(64) = 13.2$, $d = 3.31$, $p < .0001$; name effect: $t(64) = 0.349$, $d = 0.871$, $p = 0.729$).

Higher Trust ratings do appear linked to decreased privacy concerns (RQ4-G). Correlations are statistically significant with both WorriedAboutLikenesses ($r(128) = -.50$) and HardToBeAlone ($r(128) = -.35$). So Trust is a significant mediator of the effect of hypothetical familiarity on privacy concerns, but there are probably other significant ones to be discovered: at most Trust accounts for 25% of the variance (WorriedAboutLikenesses) or at least a mere 7.5% (DontMessWithMyStuff and EmbarrassedByMess).

Respondents ranked the "desk" video as most concerning (votes: most 45—12—6—2 least), followed by "bedroom" (most 14—43—6—2 least), then "photo" (most 4—4—33—24 least), then "gaze" (most 2—6—20—37 least) (RQ4-H). It makes sense with our gut intuition because "desk" and "bedroom" are more invasive, we think, than mere "gaze" or touching a "photo."

5.8 Discussion

5.8.1 Implications for Design

All four studies show a large effect of framing the relationship of the local user with the robot operator. This suggests that robot designers should think about how frames like these could be encouraged via robot appearance, what the robot says, and how the robot is advertised. Designers should also consider how different contexts and cultural factors, like the popular media and even other people, could impose unwanted frames over a user's interaction with a robot.

We are beginning to understand some details of how framing works. We see, starting in Study 2, that people can be prompted to switch between interpretive frames relatively easily. Some frames may "stick" better than others, though, even changing the way framing information is interpreted. Study 4 suggests that certain privacy concerns decrease as time passes or as people experience different frames. We believe people simply get used to the concerning behaviors and lower their guards, although an alternative explanation is that experimental realism begins to decrease with the second framing. An understanding of how privacy concerns about robots change with long-term usage is crucial; if concerns wear off or change in type after a few hours or days then robots will need to transition smoothly between these two phases.

5.8.2 On Methodology

It is surprising that the framing effect was large even with *no* signs of the operator's presence beyond the framing text. It was even present (and large) between subjects, when there were no hints about a contrast between the two frames. We would hypothesize that adding signs of the operator's presence would prolong or even increase framing effects as long as they agree with the frame. On the other hand, we are interested in what happens when conflicting information about the frame is presented to an observer, e.g., if the observer is told the robot is being used to inspect a leaky pipe but instead starts picking up your personal items.

We chose a methodology that uses animated video stimuli, quantitative data, and quick, online recruitment. We believe the animated videos are useful for studying experiences that are difficult to produce in the laboratory, which includes anything from natural robot gestures to scenarios in outer space! Also, shorter studies can operate like pilot studies when aspects of the methodology are not well-established, helping you fix problems before wasting months on a large, one-off experiment. Similarly, taking quantitative data yields early effect size estimates so we can choose proper sample sizes.

Some qualitative methods can also be short and easy to iterate on, such as focus groups. Showing our framings and videos to a focus group would be another way to explore which variables are important. Although we would not get effect size estimates or the type of insight into relationships between variables offered by PCA, a focus group conversation could explore a much broader selection of

concerns and use nuanced follow-up questions to greatly increase our confidence that participants' responses mean what we think they mean.

5.8.3 Future Work

Future work should concentrate on which frames have the largest effect sizes on privacy concerns. We have only looked at the difference between a familiar person and a stranger operating the robot in these first 4 studies. Here are some other aspects of the frame one could manipulate: the level of control the operator has over the robot's action (full teleoperation vs. supervision); how invasive the operator's actions are expected to be based on his/her role (police officer vs. tourist); morphology of the robot; robot nonverbal behavior (e.g., smoothness of movements, posture); or the operator's social presence (via a face on a screen or operator name tag). Knowing which aspects of the context most influence user perception of privacy risks (as in Hancock et al. [97]) will help researchers and robot manufacturers know what to focus on.

The framing effect itself should also be studied. How is the frame encoded and then used to interpret subsequent information? How much longer does it last beyond the four videos and 22 survey items in Study 4? What happens, exactly, when people switch framings? Also, studying the individual differences (e.g., personality traits) that moderate framing effects would help identify groups that are especially sensitive.

We are motivated by privacy. Understanding framing will help us avoid privacy

violations. More research could help us discover how privacy judgments are different in HRI; we want to identify what type(s) of privacy exactly will be problematic and which people to target with our solutions.

5.9 Lessons Learned

The findings presented in this chapter suggest that privacy concerns are not assessed on the raw actions people take (e.g., making eye contact with you or opening the drawer of your desk) but on the user's *interpretations* of these actions. For example, if the robot makes eye contact with the user, it could be interpreted as friendly interest, a creepy stare, or a variety of other things depending on the context. The level of privacy concern might therefore be quite difficult to predict using just the objective facts of the robot's actions, the objects that are involved, and other details about the surrounding environment.

This suggests that people sometimes have privacy concerns about activities that require higher-level interpretation to identify. Consider a few examples that are especially laden with privacy implications: "sneaking," "spying," "eavesdropping," "[computer] hacking," and "stalking." Each of these contains more interpreted meaning than other, more neutral words for the same sort of activity. For example, compare "eavesdropping" to just "listening," or "stalking" to just "following." One can imagine a user's expression of a privacy preference: "I don't mind if the robot looks at me, as long as it doesn't *stare*," or, "I want the robot to *pay attention* to me, but not to *stare* at me." To predict which of these higher-level interpretations

will be made by users, a privacy-sensitive robot would need to know the frame within which they are interpreting their experiences. It could be difficult for a robot to detect a user's current frame, or to model all the possible frames and what causes transitions between them. If it did these things, however, it could test or even learn techniques for influencing the current frame itself.

5.10 More on Measuring Privacy

"Privacy" can be difficult to measure because it can be used in multiple different senses in English, and even when a single definition is agreed upon (see Section 7.2.7 for some scholarly attempts) it remains multidimensional. This section is about my experiences trying to measure privacy concepts (or "constructs") for my second and third contributions. My purpose is to give privacy-specific advice to others who follow in my path.

5.10.1 Choosing what to measure

It seems unlikely that many researchers will want to measure nonspecific "privacy concern" or other attitudes about "privacy" in general, where "privacy" is taken to include *everything* in the taxonomy we presented in Section 2.1.3—everything from protecting your personal information to how close the robot stands. Instead, it will usually be appropriate to choose from among the lower-level facets of privacy. The first choice you will have to make is which facet(s) you care about, and at what

level of abstraction. This latter consideration is important: do you want to measure information privacy in general, for example, or just the part about the *collection* of information? If you choose to measure *all* of a construct that is multidimensional (like information privacy concern), be sure that your measurement instrument covers all of its component dimensions. You will also need to decide whether to report your results as a single, combined construct or several distinct ones.

It is important to be clear about what *target* you are asking participants to describe or rate in a self-report measure. There are a lot of different things to ask about in human-robot interaction research. You could ask people to respond to something about this robot, this model of robot (e.g., Roombas), this class of robots (e.g., vacuum cleaning robots), or robots in general. If you are asking about the robot they are interacting with, you have further choices to make: besides asking about the robot in general, you can also ask about a certain feature of behavior of the robot, a certain part of an interaction, or an entire interaction. Finally, you could ask a participant about him- or herself. The target that you are asking your respondents about should be clear, both to the respondents and to the readers of your published reports.

It is also important to be clear whether you want to measure a "dispositional" construct or a "situational" construct. A dispositional construct is a relatively stable "trait", whereas a situational construct is more like a "state" that varies more with time and circumstance. If you are measuring someone's concern about being seen by the robot while changing clothes, for example, you should be clear about whether you mean their level of concern right now (or averaged over the

duration of an interaction) or the extent to which they are the sort of person who usually feels worried or shy about changing in front of others. This distinction is important both for questions about the respondent (how they feel right now vs. their predisposition towards feeling that way) and about the robot (how the robot was in this interaction vs. how it is in general).

Once you have clearly specified your construct in these ways, make sure you can also articulate an *operational* definition—i.e., a description of what phenomena you would observe at different levels of the construct—before proceeding to measurement. For example, heightened concern about surveillance might cause a person to glance at the robot more often and say they felt like it was "watching" or even "monitoring" them. If you don't know what would be *observably* different about a person at different levels of a construct then you can't measure it.

5.10.2 Choosing a measurement instrument (or several)

There are several types of measurement to choose from. Self-report measures are perhaps the most popular in HRI, and work well for internal states (e.g., attitudes, desires, thoughts, and feelings) that are hard to observe externally—but only if participants can describe them clearly and accurately. See Furr and Bacharach [81] for an introduction to measurement theory ("psychometrics") and especially self-report measures like scales and interviews. If participants cannot be trusted to give unbiased responses, researchers could use physiological measures of things like heart rate, galvanic skin response (GSR), and eye gaze [242]. You could also

design the study scenario such that a certain, observable behavior is unambiguously indicative of the target construct, but also so that the participant is unaware of this and therefore largely incapable of "faking it." These carefully designed observational measures as well as physiological measures can be especially useful when *mentioning* privacy in an interview or questionnaire could make respondents more worried about privacy than they might have been otherwise.

5.10.3 Validating your measurement instrument(s)

Validation of measurement instruments—i.e., making sure they are working properly—is always an important step [81, 60]. *Validity* should be thought of as a property of the way researchers interpret the output of an instrument: "validity is 'the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses' of a test" [81]. So validating a measure of (supposedly) information privacy concern, for example, helps us decide whether it is valid to interpret a certain reading or score from that measure as indicating a certain level of information privacy concern and not something else.

There are multiple facets of validity—Furr and Bacharach [81] list five (see Ch. 8)—that are each measured in different ways. A simple check is to compare your measure with other measures of the same construct—they should be strongly correlated. Since privacy is so complex and multidimensional, however, it will also be very important to make sure you are measuring the right privacy construct and not a related one. To do this you will need a taxonomy that shows the

hierarchical relationships between privacy constructs like the one we present in Section 2.1.3. You can then estimate the validity of your measure as follows. Conceptually, a particular privacy construct in the taxonomy is more similar to its "parent" construct than to other constructs at that level of abstraction (i.e., "aunt/uncle" constructs), and it should also be distinguishable from constructs at its own level with the same "parent" (i.e., "brother/sister" constructs). We can estimate the strengths of these relationships by inspecting the correlations between measures of the different variables. For example, to validate a measure of concern about personal space you might check that it correlates more strongly with general measures of physical privacy concern than with general measures of information, psychological, or social privacy concern; you could also check that it is not too strongly correlated with measures of a sibling construct like concern about territory, for example. In more technical terms, these checks help you estimate the *convergent and discriminant validity* of your measure [81].

5.10.4 Common mistakes to avoid

We have learned several other lessons over the last few years about measuring privacy constructs:

- **Don't place any special confidence in a "validated" scale if you're going to change it or use it in a new way.** Validated scales exist to measure related constructs about, e.g., Internet privacy. These scales might have useful items that you could borrow and modify to fit your purposes.

Even if you use the entire scale and only change the word "Internet" to "robot", however, you can no longer make strong statements about the scale's performance based on previous validation studies. In terms of your confidence that the scale measures what you want it to measure, think of it as starting from scratch.

If no existing scale measures exactly what you want to measure, then it's usually better to write your own (taking inspiration from existing scales is recommended, though). Just make sure to adequately test a new scale before using it, and also to monitor its performance in each study to make sure it works as you intended.

- **Be careful not to create or inflate privacy concerns by asking about them.** It can be difficult to measure someone's current level of privacy concern without changing it. Some people might not have thought at all about privacy until you mention it to them in an interview or survey question. Be aware that this could influence their responses to that question as well as future questions. You might do this intentionally, however, if it helps you answer your research question—e.g., bringing up privacy to brainstorm solutions with a focus group.

- **Don't use words like "privacy" that can have multiple meanings without making sure you know which meaning people are responding to.** One way to do this is to specify what you mean to try to control how people interpret your prompts; you might instead leave it ambiguous, how-

ever, if you *want* to get responses to multiple different interpretations. Either way, you should try to measure how each person interpreted your words so you can correctly interpret their responses. In an interview you might ask follow-up questions, or in a scale you might make sure the responses are unimodal and strongly correlated with the other items.

5.11 Choosing our Next Contribution

For our next (fourth) contribution we had several methodological goals for getting beyond the limitations of our first two studies. First, we wanted to do a longer-term study to observe participant behaviors after they get used to the robot—i.e., to avoid only recording novelty effects. Second, we wanted to venture outside a controlled, laboratory (or online) setting to places where people already spend time. Lastly, we knew that self-report measures like questionnaires come with the risk of response biases such as answering so as to please the experimenter (i.e., in response to "experimenter demand") or to be "socially acceptable." We decided to protect against these biases by including observations of participant behaviors in our measures for the next study.

For our construct of interest we chose mental models—i.e., the description of a robot that a user builds up in their mind. Inspired by the findings of Lee et al. [142], we were especially interested in how people figure out whether a robot can record video or audio when cues are ambiguous or absent. We hypothesized that understanding this is fundamental to predicting and manipulating people's privacy

concerns about interacting with the robot. For example, robot designers might be able to predict that users will ignore or forget about a certain sensor; they could then choose to make it more obvious to help users make better-informed decisions about their privacy. By studying multiple interactions between the same person and robot over several weeks, this next (fourth) contribution will also speak to how designers might need to address privacy concerns differently at different stages of the user-robot relationship.

6 Forming a Mental Model of the Mobile Shoe Rack: a Long-term, Qualitative, in-the-Wild Study

The work presented in this chapter had not been published at the time of writing, but it was not done alone. Jeffrey Klow served as the "wizard" and did most of the equipment setup. Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, and Madison Browning helped to design and run the study, as well as to manage all the data. Frank Bernieri, Cindy Grimm, and Bill Smart supervised my work and served as mentors (as they did for the other two experimental contributions, too).

Abstract– Most people don't have direct access to knowledge about the inner workings of robots—instead, they must develop mental models to explain and predict robot behavior. Despite being at the core of how people understand robots, this process of forming mental models is not well-understood. We report findings from a long-term, in-the-wild, qualitative, hypothesis-generating study that was designed to identify some characteristics of the mental model formation process for further research. Participants of diverse ages had multiple interactions with the robot over six weeks in a non-laboratory setting. A novel, non-anthropomorphic robot was created for the study with a realistic use case: storing people's shoes while they are in yoga class. The robot's behaviors were varied systematically to study how people would account for abrupt changes.

This paper reports findings from a case study analysis of 28 interviews conducted over six weeks with six participants. These findings are organized into five themes: (1) variability in duration of mental model development, (2) types of reasoning and hypothesizing about the robot, (3) borrowing from existing mental models and use of imagination, (3) attributing sensing capabilities where there are no visible sensors, (4) judgments about whether the robot is autonomous or teleoperated, and (5) experimenting with the robot. Specific suggestions for future research are given throughout. This work demonstrates the fruitfulness of long-term, in-the-wild studies of HRI, and we provide recommendations for making them more efficient and focused. It also represents an early study of mental model formation about robots, a foundational topic for understanding and designing human-robot interactions.

6.1 Introduction and Motivation

Not everybody understands robots. The people who design, build, and program a robot understand it because they have insider knowledge about its intended functions and inner workings. They know what it can sense, how it thinks, the rules governing its behavior, its learning capabilities, and its connectedness to other computer systems. Some people understand the robot because they made it, but they cannot pass this understanding on to users directly. Instead, users typically rely on what Don Norman calls the "system image" [174]: the parts of the system that are observable by users, including how it looks and behaves. Users then use

Figure 6.1: Reenactment of a typical hallway scene during the study.

these observations to develop a "mental model" of the robot that explains and predicts its behaviors [174]. This mental model can include the user's perception of what the robot can sense [142], how it processes information (including memory and learning) [69], how it behaves (including its "personality", if applicable) [259], its role in its setting [162], what it knows [143], and other abilities and attributes [48].

Understanding how people form mental models of robots is critical for a number of applications. When working with robotic teammates, for example, we will need to understand what they know and what they intend to do next [124, 212]. Designers will need to understand how to make a robot and its user interface in a way that helps people figure out its sensing capabilities—this is important, for

example, for making judgments about personal privacy [209]. It will also be useful to influence users' mental models to encourage them to like the robot and accept it for long-term use [55].

It is not fully understood how people form mental models of robots. Very few published HRI studies have measured mental model formation over time, and those that do (e.g., Stubbs et al. [228]) seem to focus on quantifying the changes in the mental models instead of describing the process governing those changes. Our study focuses on the process instead. We analyze *how* people use their pre-existing mental models as well as observations from multiple interactions with a robot to develop a mental model for it over time.

This paper reports findings from analysis of 28 interviews conducted over six weeks with six of the participants who interacted with our robot, the "mobile shoe rack." An overview of our data is presented in Section 6.7. Our study took place "in the wild" [206]—specifically, in the hallway outside a yoga classroom on a university campus. Our analysis considered multiple aspects of the interviewees' mental models, including what the robot can sense and infer, its rules of behavior, and how connected it is to humans (specifically, whether it was being remote controlled by members of the study team). Our particular focus was on what they noticed about the robot and how they reasoned about these observations and used them to update their mental models. To capture this process we did a qualitative analysis of the interviews to extract chronological descriptions of the interviewees' thought processes. Each interviewee's responses were first analyzed

as a case study; we then performed a cross-case analysis to identify themes using all six interviewees' responses.

Findings are organized into five themes in Section 6.8, including how people reason and hypothesize about the robot's behaviors, borrow from existing mental models and use their imagination, and judge whether the robot is autonomous or teleoperated. Our report contributes early documentation of phenomena from our study that we recommend as high priorities for HRI research. We also contribute in Sections 6.9 and 6.10 new questions about mental model development that were generated as part of the data analysis process. Besides providing data-driven research questions for researchers, this work should also appeal to designers as a key case study for grounding the design of robots and their behaviors in the experiences of real users in realistic use cases. We hope to help designers better understand how to shape a user's mental model through the design of the robot's appearance and behavior.

6.2 Background: Mental Models

"Humans construct internal representations, or models, of objects in the environment, such as other people, animals, and machines..." [182]. These mental models are "the conceptual frameworks that support people's predictions and coordination in a dynamic world" [125]. Mental models are often "incomplete" and "not accurate"; also, "Previous research suggests that people hold sparse, primitive mental models of unfamiliar objects, technologies or ideas with which they have little ex-

perience" [182, 87]. People then change their mental model of something based on their interactions with it to make them more useful for achieving their goals [175]. Note the difference between a mental model of a robot and a "knowledge estimate" of what it knows [124]. A mental model is also different from a "shared mental model", which is one agent's model of its team and of the team's task.

Several ways to measure someone's mental model of a robot have been documented. One strategy is to ask direct questions about the robot and its behaviors: e.g., "Describe the robot's task."; "Why does the robot stop?"; "Why do its lights flash?" [259] The language someone uses to describe the robot and its actions can also be informative. For example, saying "it is an aggressive robot" implies it has personality traits and saying "the robot is angry" implies it has emotional states, whereas "the robot hit me" implies neither [82]. It may also be possible to measure a person's experience of "cognitive conflict" or dissonance when they encounter the differences between a real robot and their preexisting notions about robots, e.g., from science fiction and the news media [148]. This tension, which Levin et al. operationalize as self-reported difficulty answering questions about robots, could indicate when people are reconsidering or changing their mental model [148].

6.3 Background: Novelty and Habituation Effects

Characteristics of the early stages of interaction with a robot do not always generalize to later interactions when people have gotten used to it. An early forewarning of this problem was documented with studies of "video telephones" at Bellcore in

the early 1990's wherein usage dropped off after a few days [74]. A clear example on a robot is documented for the case of "Valerie the roboceptionist." People were still using "Valerie" 9 months after "she" was deployed, but the duration of time each person spent with "her" was much higher during the first week [88]. The authors called this a "novelty effect"—i.e., the fact that the robot was *new* caused a temporary increase in people's interest levels [88]. Lemaignan et al. [146] have proposed a theoretical model that explains the novelty effect on anthropomorphism by how people borrow and adapt their existing mental models to understand the robot.

A related phenomenon is the "habituation effect", in which the effect of some aspect of the robot on the interaction changes as people get used to the robot. It was observed in one study that participants allowed "the robot with mechanoid appearance to approach closer than the robot with humanoid appearance" in the first interaction, but this effect faded with subsequent interactions [128]. This shows how important it can be to run a long enough study to see whether an effect remains stable over time.

6.4 Related Work in HRI

6.4.1 Mental Models of Robots

6.4.1.1 Expectations

Some of the prior work on mental models in HRI is related to people's expectations about a robot's abilities and how these expectations can be influenced by first impressions. For example, one study [143] found evidence that someone's expectation of what the robot *knows* is influenced by information about where it's from—e.g., a robot from New York City is expected to know about New York City landmarks.

Several groups have called attention to the mistakes people make when inferring robots' abilities from their actions [48, 133, 221]. One study [48] found evidence that manipulating a robot's apparent speech capabilities can change someone's perception of its physical capabilities even though the two are often unconnected in robotic systems. Richards and Smart [221] have argued that this will be especially true for humanoid robots, and early experimental results seem to agree [133].

The importance of cues people use to guess sensing capabilities was highlighted by a study of privacy concerns about "Snackbot" [142]. Very few of the people who were interviewed could identify what sorts of sensors it had—they were especially surprised to learn about the omnidirectional camera—or guessed that it was recording audio and video.

Our study also looks at the expectations people have about a robot from preexisting stereotypes or from cues given by the robot. In addition, our study is

long-term, qualitative, and conducted outside of a laboratory setting, which is rare in prior work.

6.4.1.2 Experimentation

There have been several studies of how first-time users of a robot have to figure out what it can sense and what it responds to. Three different studies report that participants experimented with the robot by touching it, talking to it, waving, or snapping their fingers [162, 177, 210]. We were wondering whether people would do this in our study, wherein we too avoid telling our participants how the robot works or making the sensors too visible.

6.4.1.3 Mental Models of Robotic Furniture

Our robot design was inspired by a series of robotic furniture designs [220, 260, 162, 227]. Mental models are at the fore for robotic furniture because there's no obvious analog—you can think of an AIBO as a dog and a Nao as a person, but how do you interact with a moving sofa, for example? Interestingly, some people have attributed sophisticated abilities robotic furniture; see, e.g., the report that "people [created] mental models of the trash barrel as having intentions and desires" [260]. The robotic trash barrel also induced some people to borrow from their existing mental models. One person, for example, treated it like a dog: "[He] called the trash barrel over by whistling and making kissing noises while waving chopstick

wrappers like a dog treat. And after he had disposed of the wrappers and the trash barrel acknowledged the trash with a wiggle, he happily noted to his colleagues, It's wagging its tail!' " Our study will be the first that follows a group of robotic furniture users over many interactions to see how their mental models of the robot develop.

6.4.2 Long-term HRI Studies

Some studies have deployed robots for longer periods of time to measure how interactions evolve. Sung et al. [231] in 2009 listed eight longitudinal HRI studies: two each in offices, schools, hospitals, and the home. Four years later, Leite et al. [145] published a survey of long-term studies of *social* robots in particular—they listed 23 total studies: five in health care and therapy, eight in education, five in work environments and public places, and five in home environments.

Some long-term studies use ethnographic methods, like the study by Fink et al. [70] wherein they placed a Roomba vacuum cleaning robot in each of nine households in Switzerland for six months. Their methods included home visits, a home tour, qualitative interviews, and cleaning diaries. They included additional activities in the interviews like Bubble Talk, the Day Reconstruction Method, drawing, and tinkering [70]. In our study we mainly used short (most were 10–20 minutes) interviews to keep things simple and minimize the burden on our participants' schedules.

In some long-term studies the robot's behaviors change over time. For example,

in one study [92] the DragonBot personalized its actions to individual children over a two-month evaluation. To keep our study more controlled, we chose not to adapt the robot's behavior to individual participants. Instead, our participants experienced two abrupt changes in the robot's behavior that were intended to resemble unannounced software updates.

6.4.3 Long-term HRI Studies about Mental Models

There are very few published reports of long-term studies that focus on changes in people's mental models of robots. One example is a study of "agent migration" cues that looks at some complex intervention scenarios [128]. Another example is the study by Stubbs et al. [228]. They did multiple interviews of museum employees who interacted with the educational robot PER every day. They seem to only perform a quantitative analysis, though, to quantify changes in participants' mental models. We instead perform a qualitative analysis to study *how* (i.e., by what *process*) people update their mental models of the robot using the things they notice during interactions. We do share with Stubbs et al. a focus on "unmediated" interactions, though, in which participants are not instructed about how to interact with the robot beforehand, nor does anybody guide them during the interaction.

6.5 Study Goals and Approach

See Table 6.1 for a summary of our research questions and study goals. This section describes them in more detail.

6.5.1 Rationale for Qualitative, Hypothesis-Generating Approach

Most HRI studies up until now seem to take a hypothesis-testing approach. They pose a particular hypothesis about certain variables—these are first operationalized and then measured to test whether the data make the hypothesis seem unlikely. This works well when you have a theory you are trying to extend (e.g., from human-human interactions to human-robot interactions) or refine.

On the other hand, mental model development in human-robot interaction is relatively unexplored, and lacks a theory to test hypotheses about. Instead of making speculative theories from behind our desks to give us something to test, we first immerse ourselves in the real-world phenomena we are studying by simply asking people for detailed descriptions of their experiences with the robot. We reason that it is better for us to let the specific research questions, variables to measure, and (eventually) theories arise from a systematic analysis of data collected in a natural setting than to try to guess them ourselves. Our approach is qualitative because we do not attempt to estimate the numeric values of any variables from our interview transcripts, as well as hypothesis-generating because our goal, besides discovering the interesting phenomena involved in this process, is to produce data-driven research questions for targeted testing in future work.

This approach has several other advantages besides being well-suited for early, exploratory research. Qualitative methods tend to focus on uncovering a *process* instead of accounting for the *variance* of certain variables. Since qualitative methods emphasize *how* two events are linked and not the magnitude of the relationship, they are well-suited for uncovering the causal chain and making recommendations for how to improve something (e.g., a human-robot interaction). Finally, qualitative methods tend to focus more on individual people and their specific contexts instead of on differences between group averages. This can help researchers identify new, unexpected phenomena.

6.5.2 Venue and robot application

We wanted to do a realistic robot deployment "in the wild" (i.e., outside the lab) for multiple weeks. We chose a classroom setting because we wanted to observe the same group of people at least once per week, and we wanted there to be relatively few absences to minimize missing data. We targeted courses for faculty and staff instead of undergraduate courses to get a wider age range of participants.

We also thought it was important for the robot to have a useful role in our chosen setting, as we have noticed that some participants (and spaceowners) are no longer happy to "play along" with the experiment when the robot does a useless task, does its task very poorly, or makes life more difficult in any other way. These participants seem to get stuck focusing on their criticisms and become resentful towards or dismissive of the robot. Instead of this, we wanted our participants to

RESEARCH QUESTIONS
• How do people form mental models about robots over time?
○ What are their expectations before their first interaction with the robot?
○ What are their preliminary models after the first one or two meetings?
○ How do they change these models over time as they experience more?
○ How do they cope with sudden changes to the robot's rules of behavior?

STUDY GOALS …	… and the methods we chose to achieve them
• Methodology	
○ Collect periodic descriptions of mental models without biasing future measurements	○ nonintrusive video ○ indirect interview questions
○ Report should *describe a process*, not *quantify effects*	○ qualitative (case study) analysis
• Robot	
○ Non-anthropomorphic robot with a minimum of sensor cues	• a "mobile shoe rack" • partially hidden webcam
○ Participants form their own mental models	• minimal instructions or description of the robot by the study team
○ Realistic use case so people play along	• storing shoes during yoga class
○ Full control of (changes in) robot behaviors	• Wizard of Oz technique
• Setting	
○ "In the Wild"	• in the hallway outside a yoga classroom
○ The same people come there regularly	• participants had registered for a biweekly class
○ Wide age range of adults	• university faculty and staff

Table 6.1: Summary of our research questions and study goals. Each goal is matched with the aspect of our study methods that fulfills it. See Sections 6.5 and 6.6 for details.

be relatively willing to interact with the robot so that their mental models would develop over time.

6.5.3 Robot appearance and behavior

We wanted the robot to have a novel appearance and function so people would not be able to just apply an existing model they had for something else. Instead, we wanted people to struggle a little and undergo a more lengthy and complex learning process. We also wanted to minimize obvious clues about the robot's sensing abilities and connectedness to humans. We avoided anthropomorphic cues (e.g., eyes, ears, a head) and sensors (cameras, microphones) that are visible and easy to identify. We also kept the robot wireless so it's ambiguous whether it's being remote controlled or sending sensor data to a human. We wanted to avoid making things obvious so we could measure how people used their preconceptions and made inferences from their observations of the robot's behaviors.

These behaviors were designed to be deterministic and not stochastic so the robot always did the same thing given the same situation. This was intended to make the robot seem more "robotic" and to cover our tracks so people would be less likely to suspect that the robot was not autonomous. In reality we were remotely controlling the robot using the Wizard of Oz technique to keep the robot's behaviors very consistent (i.e., no unexpected errors) and deal more intelligently with unanticipated circumstances. We designed the robot's behaviors to be realistic for a present-day autonomous robot.

6.5.4 Measurement Goals

We chose to use semi-structured interviews for two reasons. First, we wanted a method based on self-reports so we could ask detailed questions about mental phenomena like thoughts and beliefs. This would be difficult to extract from simply observing participant behavior, for example. Second, we chose semi-structured interviews instead of (fully structured) questionnaires so we could ask follow-up questions to interesting responses.

We avoided asking people direct questions about the robot's capabilities because we did not want them to pay an unnatural amount of attention to the robot, or to be artificially motivated to try to figure out how it worked. If some participants interact with the robot multiple times without ever considering whether it can hear, for example, we do not want to spoil our ability to record that phenomenon by asking about hearing earlier in the study. Instead, we chose general questions about their interactions with the robot—what they did, saw, heard, thought, and felt, as well as their impressions of it.

We were also careful not to suggest certain answers to the interviewees or lead them down a certain mental path. Instead, we asked open-ended questions like, "Tell me about your experiences with the mobile shoe rack today" and to use the interviewee's words when asking for more details: "You mentioned that it has a 'motion sensor'—could you talk more about that?"

6.6 Methods

We deployed a "mobile shoe rack" in the hallway outside of a yoga classroom on a university campus during class time.

6.6.1 Venue and Population

The yoga class we chose is attended by faculty, staff, and graduate students. People who attend must remove their shoes before entering the classroom, but there is no good place to put their shoes—they cause a mess in the classroom and are not allowed in the hallway. A robotic shoe rack gives people an approved place to leave their shoes in the hallway, and can even drive up to people to collect them. Its movement was restricted to a taped-off area around a bench where people take off their shoes (see Figure 6.1). It turned out that students rarely talk to each other in this setting, which suggests that our focus on individual instead of group phenomena was a good fit.

Participants were recruited for interviews by e-mail and in person. Also, anybody who stepped inside the taped-off area was captured on video and audio, as explained by signs posted on the walls. This second group of participants included some students from other classes happening before or after our target class. Our class ran from 12 to 12:50pm on Mondays and Wednesdays.

6.6.2 Video and Audio Recording

A video camera was mounted high up on the wall opposite the taped-off bench area where the study took place. The camera was positioned so its field of view ended at the tape border and extended high enough to see participants' faces (see Figure 6.1). The camera was configured to record for the entire time when the mobile shoe rack was inside the taped-off area. The camera was pointed up when it was not recording so that it could not see anyone.

6.6.3 The Robot

6.6.3.1 Appearance

The mobile shoe rack was built on the Turtlebot 2 platform. Figure 6.2 shows how it looked to participants. The top surface of each shelf was covered by paper with pale orange shoeprints painted on it to suggest that it's intended for shoes. Besides this, the robot had no nametag or instructions posted on it. It was identified as a shoe rack, however, by two signs at the edges of the taped-off area as well as during an announcement by the author (i.e., Matthew Rueben) in the yoga class and in e-mails to people with offices in the hallway.

6.6.3.2 Networking

The wizard's computer was connected to the robot's netbook over 5 GHz Wi-Fi. We had network connection dropouts—short (approximately 1-15s) times when the wizard either couldn't drive or couldn't see any video from the robot. During these times the robot stopped. This was most conspicuous in Condition #1 (see Section 6.6.4 below) when the robot happened to be in the act of driving across the taped-off area or turning around. It was less conspicuous when the robot was in the home position in Conditions #2 and #3 because the robot only moved occasionally in that situation. Sometimes this happened as often as once per minute or two.

6.6.3.3 Sensors

A webcam with a built-in microphone was mounted underneath the front edge of the shoe rack, pointing forward to give the wizard a video feed to drive by. Audio was recorded from the microphone, but not piped to the wizard. We intended to be able to hear people talking about the robot using this microphone, but there was a lot of noise from the robot's motors, wheels, and from people walking on the creaky floors that made it very difficult to hear most conversations. We could, however, hear the robot's beeping noises in the recordings to confirm that the wizard had triggered them at the right times.

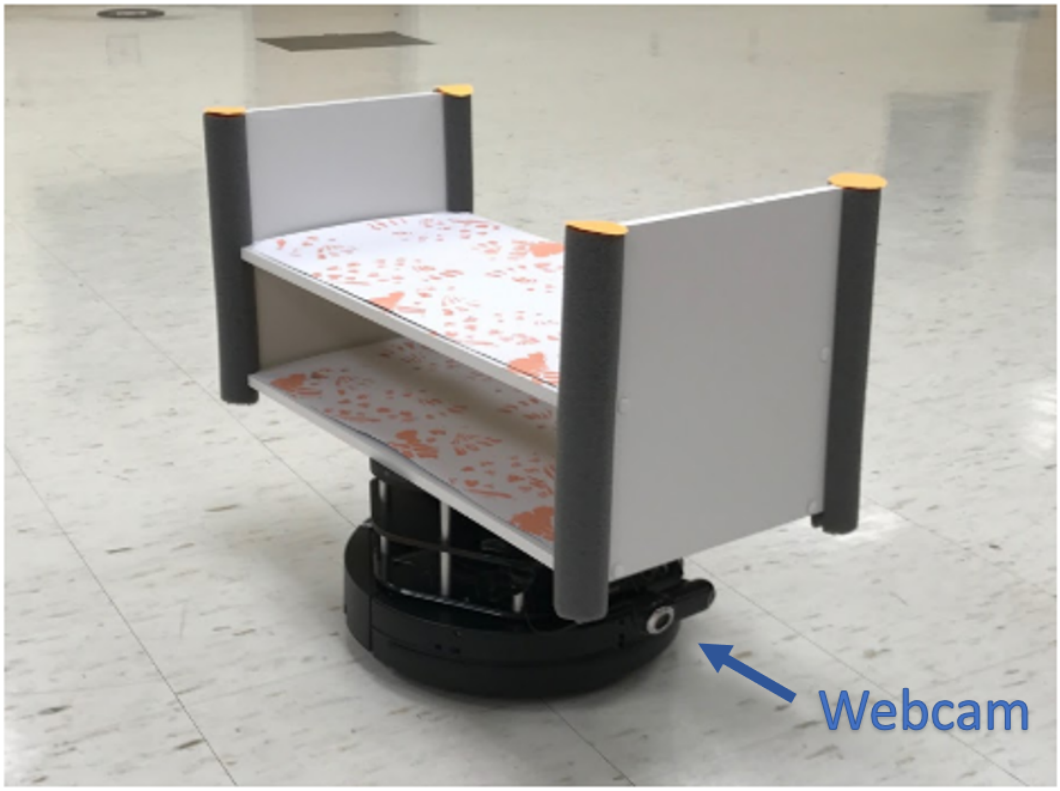

Figure 6.2: Closeup of the "mobile shoe rack" robot. Webcam is visible under the right edge of the shoe rack. Foam was wrapped around the four corners of the shoe rack to prevent hurting people if the robot bumped into them (or vice versa) even though the robot drives pretty slowly. An orange circle capped each foam tube to make it possible to automatically track the robot's pose from the wall-mounted webcam.

6.6.4 (Simulated) Robot Capability conditions

The robot's behaviors were manipulated to produce three conditions. Each condition consisted of a set of behavioral rules that were designed to simulate a set of sensing and processing capabilities. During each condition the robot was controlled by a human wizard according to that condition's behavioral rules. Each condition lasted for two weeks. The robot capabilities for each condition consisted of the capabilities from the previous condition plus some new ones. All sounds made by the robot were custom-made by the study team using the free software Audacity. See Figure 6.1 for an overview of the study area.

6.6.4.1 Condition #1

- **Abilities**: The robot can detect obstacles. It can also detect people as different than inanimate objects.
- **Sounds**: We created a "bee-boop" sound to get people's attention and signal that the mobile shoe rack is ready to receive (or return) their shoes. The two tones were a higher pitch followed by a lower pitch in quick succession. The "car horn" sound was based on the sound made by an impatient driver on a car horn: two medium-length blasts in quick succession.
- **Behaviors**: The robot drives repeatedly along the length of the bench. It stops and makes the "bee-boop" sound when someone is on or right in front of the bench. It also stops for obstacles, and behaves differently for human

obstacles than for inanimate objects. The robot would steer around inanimate objects using clunky 90-degree turns. Humans received the "car horn" sound—if the person doesn't move after a few seconds, the sound is given again. If this also fails, the robot would steer around the person like it does for inanimate objects.

6.6.4.2 Condition #2

- **Abilities**: The robot can do everything from Condition #1, plus distinguishing between individual people and remembering who it has visited within the entire taped-off area. It can also tell when shoes are added to or removed from the rack, but only after it makes the "bee-boop" sound and is waiting for an interaction.

- **Sounds**: The "sad sound" was designed to make the mobile shoe rack sound disappointed. Two tones were used as for the "bee-boop" sound, but these were longer, and each fell in pitch.

- **Behaviors**: Starting in this condition, the robot does not waste time driving back and forth when there are no people to serve; instead, it waits in a "home position" in the middle of the hallway near the edge of the box that's farther away from the yoga classroom. The robot's camera side faced down the hall towards the classroom, watching for people to enter the taped-off area. Every 10 seconds or so, the robot turned slightly to one side and then to the other to

look for people. If the wizard saw through the robot's webcam that someone enter the taped-off area at any point, he chose that person as his target and began driving straight towards them—this was the first condition in which the robot drove diagonally. If the robot's path to its target was blocked by anything, whether person or object, it would stop and drive around it without honking and with smoother maneuvers than in Condition #1. If someone stepped right in front of the robot and caused it to stop suddenly, however, it gave the "car horn" sound. The robot also issued the "sad sound" if it reached its target, issued the "bee-boop" sound, and waited several seconds without experiencing any exchange of shoes.

6.6.4.3 Condition #3

- **Abilities**: The robot can do everything from Condition #2, plus detecting when someone is making eye contact with the robot. It can also detect when its current target is in the process of taking off their shoes.

- **Sounds**: The "blabber" function was designed to make it sound like the shoe rack was talking, albeit in a robot language of electronic beeps. We created a library of 12 audio clips ranging from 0.4-1.1s long that each sounded like a single word or short phrase; each "blabber" action by the robot would randomly play one of these clips. The beeps varied in length and timing to mimic the cadence of human speech.

- **Behaviors**: The robot mostly behaves the same as in Condition #2 except that as it is driving towards a target person, it does a "blabber" if that person looks at it. After a cooldown period of about 5 seconds, the robot will "blabber" again if the person glances at it, or has been staring at it. Hence, if someone stares at the robot while it is engaged with him or her, it will "blabber" once every 5 seconds. In this condition if the robot comes up to someone while they are taking off their shoes it will wait until they finish and place their shoes on the shoe rack before looking for another person to approach. The robot also follows a smoother trajectory instead of always stopping to turn in place.

6.6.5 The Wizard

The mobile shoe rack was remotely controlled by Jeffrey Klow, another graduate student on our study team. He practiced our robot behaviors at the chosen venue for four weeks interacting with students from two different classes before the start of the main study. He became quite consistent, although he still deviated from the behavioral protocol sometimes. This included: (1) driving closer to or farther from the bench in Condition #1, (2) forgetting to play the correct sounds at the correct times, especially when there was a lot going on like in the more advanced conditions; and (3) imprecise timings—e.g., for the 10 seconds between looking to either side in the latter two conditions, or for the 5-second "blabber" cooldown.

Error 2 in particular probably made it harder for those participants who saw it to figure out the patterns in the robot's behavior.

6.6.6 The Wizarding Interface

The robot and wizard were both running the Indigo release of the Robot Operating System (ROS [186]). The wizard's driving interface was a custom version of the keyboard teleoperation interface available in the *turtlebot_teleop* package in ROS. The only changes were the addition of keys that trigger the robot's four different beeping behaviors: the bee-boop, sad sound, car horn, and blabber. The wizard had access to two camera views: one from the forward-facing webcam on the Turtlebot and the other from the wall-mounted webcam that recorded participants' behaviors.

6.6.7 A Typical Class Time (Procedure)

The author carried the mobile shoe rack to the venue and placed it in the "parked" position at around 11:35am. This experimenter was sure to leave quickly without talking with anyone. The robot then began driving at around 11:40 or 11:45am by driving away from its "parked" position against the wall and beginning the behavioral protocol for whichever condition was active during that two-week period. The previous class was scheduled to end a few minutes later at 11:50am, and our class began at 12 noon. The robot continued to drive until 12:05pm, 5 minutes

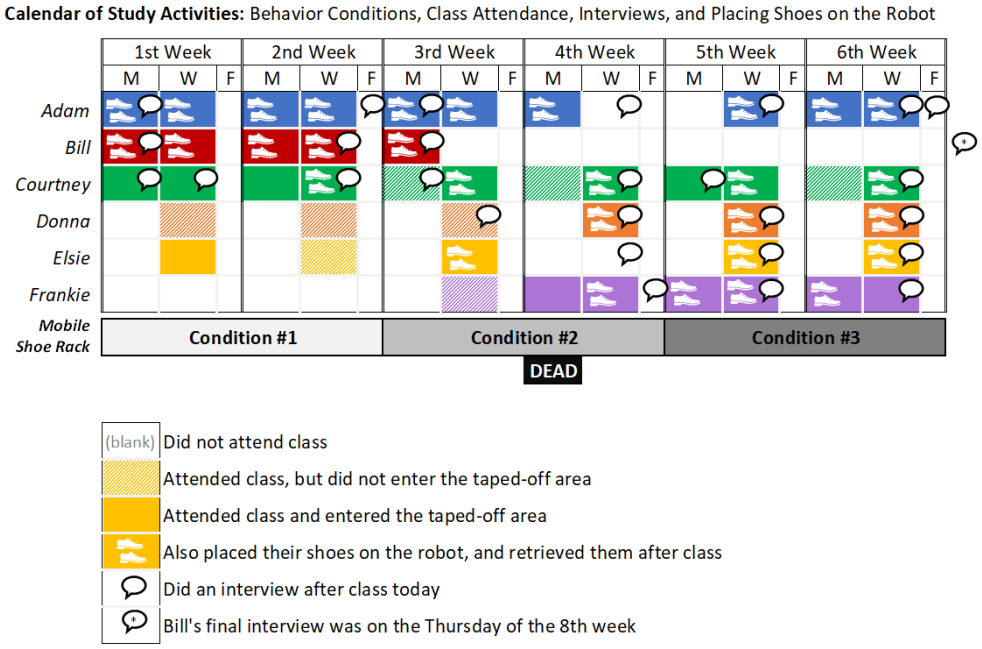

Table 6.2: Study schedule and diagram of participant attendance and participation (i.e., whether they entered the taped-off area or put their shoes on the robot), as well as when the interviews happened and when the robot conditions changed. On the 4th Monday the robot was out of battery and is labelled as "DEAD!" (see Section 6.6.8). Bill and Frankie were the instructors, and traded places midway through the 3rd week.

after class began, to catch any students who came late. The robot then parked until 12:40pm, 10 minutes before our class got out, but video and audio were still being recorded from the wall-mounted camera in case people interacted with the robot while it was parked. The robot then began driving again to return people's shoes using the exact same behaviors as before class. When the robot parked itself at about 1:05pm, an experimenter came up to collect the robot.

6.6.8 Timeline of Study Activities

Table 6.2 shows the study schedule, including the three behavioral conditions and the dates of all the interviews with each of the six participants who volunteered to be interviewed.

Study activities began with an announcement by the author during the first yoga class of the term. Students were told to expect the "mobile shoe rack" before and after class, and that they could put their shoes on it if they wanted to. No study activities happened that Wednesday; trials started the following Monday and spanned six weeks (12 class times) total. Classes continued after the end of the study for three more weeks before the term was over.

These six weeks were divided evenly into three conditions of two weeks each. Transitions between conditions were done all at once and without warning the students—on the first day of a new condition, the robot starts doing its new behaviors as soon as it starts driving.

One day was an exception; Monday of Week 5, during Condition #2, the robot was accidentally left unplugged overnight and had a dead battery, so it was placed in its "home position" (see condition details above) and left there, now an "immobile shoe rack", for the entire time that the robot was usually in the taped-off area.

Interviews were done about once per week. We targeted days when we thought interviewees would have more to say: the first day of a new condition when the robot's behavior has suddenly changed, and the last day of a condition when par-

ticipants have had two weeks to observe it. Interviews did not always conform to this schedule, though, especially for the two participants we named "Elsie" and "Donna" (see Section 6.6.12 below) who only attend class—and only did interviews—on Wednesdays.

6.6.9 Interviewee Questionnaire

All interviewees completed a questionnaire to help us gauge the diversity of our sample. The questionnaire begins with demographics: age, gender, race/ethnicity, major (if they attended any college), year in school (for graduate students), department (for faculty), and education level (for faculty and staff). The rest of the questionnaire is about technology usage and expertise, as well as attentiveness to one's surroundings. The items were intended to measure openness to adopting new technologies, experience with computers and robots, interest in robots, cell phone and social media usage, ability to figure out how things work, and tendency to notice details around oneself.

6.6.10 Interview protocol

The author did all the interviews except for several that were done by study team member Eric Zimmerman. Interviews were recorded on a handheld audio recorder, and interviewees were compensated $10 at the end of each interview for their participation.

Interviews were semi-structured: interviewers were given a list of carefully-designed questions to ask, but were also trained to give encouraging prompts and ask follow-up questions. The full interview guide is included in Appendix A at the end of this dissertation.

The first part of the interview was meant to get the interviewee's account of everything they noticed about the mobile shoe rack since the previous interview. We wanted specific details, so we asked follow-up questions to get the exact sequence of actions done by both the robot and the interviewee.

The second part of the interview asked more general questions. Some examples include, "What is your overall opinion of the mobile shoe rack?"; "If you could make it better, what would you change?"; "Describe all the shoe rack's behaviors as best you can"; and, "Would you want the mobile shoe rack in your home? Why or why not?" We also asked, "Do you have any questions about the mobile shoe rack?", but the interviewer did not answer these until after the final interview.

At the first interview, the interviewer also asked about the expectations that the interviewee had about the mobile shoe rack after first hearing about it, and how these expectations compared to their first impression when they actually saw it.

After the last interview, the interviewer asked direct questions about the interviewee's mental model of the mobile shoe rack. The questions were specific and covered a wide variety of capabilities: "Can it see? How far? Can it distinguish people from inanimate objects?"; "Can it hear?"; "Does it remember that you've already put your shoes on it?"; and several others.

After these direct questions, the interviewer revealed that the study was focused on how they figured out the robot's capabilities. A description was given of what the robot's capabilities were in each condition. It was also revealed that the robot was driven remotely by a human. Interviewees could react and ask any questions they had about the study.

6.6.11 Interview Analysis

All the transcription and analysis of the interview recordings were done by the author. Each interviewee's experiences were analyzed as a separate case study. Interview responses were placed in chronological order to form a coherent story for each interviewee. The analysis focused on how each interviewee used their actions and observations to form a mental model of how the robot works. The case studies were then reviewed all together in a cross-case analysis to identify emergent themes. Whenever an interviewee's response was unclear or seemed to contradict the robot's typical behaviors the video recordings were reviewed to find out what really happened.

6.6.12 The Interviewees

All interviewees were recruited from the faculty-staff fitness yoga class that the mobile shoe rack was serving. Six people volunteered to interview with the study team. We have changed their names for confidentiality. "Adam" is a male in his

30s who does STEM outreach education and professional development for teachers. "Bill" is a male in his 60s and the main instructor for the faculty-staff yoga class, but left for vacation after the 3rd Monday of the study. "Courtney" is a female graduate student in her early 20s studying marine resource management. "Donna" is a female in her 70s who was faculty in the College of Pharmacy and then worked at Student Health Services—she only attends class on Wednesdays. "Elsie" is a female in her 40s who works as a licensed psychologist for Counseling and Psychological Services—she also only comes on Wednesdays. "Frankie" was the substitute yoga instructor for "Bill", so her first day was the 3rd Wednesday of the study; she identifies as gender non-binary and is in her late 20s. Table 6.2 shows which days each interviewee attended class, and when the interviews actually happened.

6.7 Overview of Interview Data

This section is an introduction to the relatively large set of interview data collected during this study. We present some statistics of what was collected and then describe some characteristics of the interviewees' responses that will set the stage for the findings presented in the next section.

6.7.1 Interview Data Collected

We did 28 interviews in total (see Table 6.2). Adam and Courtney each did seven since they were present for the whole study; Donna (who did four) and Elsie (who did three) volunteered partway through. Bill did three before leaving on vacation and one after returning (and after the study ended); Frankie, who replaced him, did three interviews. All but the final interviews lasted around 5–20 minutes except for one with Adam that was almost 30 minutes. The final interviews lasted longer because of the direct questions about mental models as well as the debrief; most were around 30–40 minutes, although Adam's was over an hour long.

6.7.2 General Observations about What People Said

6.7.2.1 Interviewees learned to focus on the robot's behaviors

For the first one or two interviews the types of responses were more varied, apparently because different interviewees had different beliefs about the purpose of the experiment. Some, like Frankie, focused on evaluating the mobile shoe rack's design and performance, giving suggestions for improvement. Others, like Elsie, focused more on their subjective impression of the robot, calling it "cute" and "fun." By the end of the study, though, the interviewees seemed to have noticed that the interviewer was most interested in descriptions of what the robot did and why. On the 4th Wednesday Adam even began his interview on a topic we wanted

without being prompted at all: “The shoe rack, interestingly enough, on Monday, was not on.”

6.7.2.2 Use of uncertain language

We noticed during transcription of the audio files that the interviewees used a lot of disclaimers like “I think” and “maybe.” In context, these did not always seem to indicate true uncertainty, but rather something like humility or careful skepticism. Interviewees rarely made statements with certainty; this could have reflected a real uncertainty, but they also could have been intentionally trying not to seem overconfident, perhaps so as to save face if their answers turned out to be wrong. Some displays of uncertainty we took more seriously, especially when an interviewee took a stance on something rather confidently and then, realizing some flaw in their reasoning, questioned themselves or even changed their position.

6.7.2.3 Interviewees did not observe the same robot behaviors

It also became clear during the interviews that each interviewee had a unique set of experiences with the mobile shoe rack. Several factors contributed to this. First, the robot’s behaviors could change from day to day within a condition because it reacted to people around it. For example, in Condition #1 someone needs to be sitting on the bench for the robot to stop and beep at them; if nobody is on the bench, the robot will never stop or make the “bee-boop” sound. Second, intervie-

wees did not always attend class, and spent different amounts of time observing or interacting with the robot (see Table 6.2). Some interviewees would walk quickly past the taped-off area and barely glance at the robot; others would pause just long enough to take off their shoes. For those who put their shoes on the mobile shoe rack, some would sit on the bench watching the robot for a few minutes whereas others would rush into the classroom. Bill and Frankie often arrived very early and others sometimes arrived late—in both cases, they were often the only person in the hallway with the robot. Other interviewees arrived during the transition time between classes when the hallway was crowded; they saw the robot interact with several other people.

This second factor—the amount of time and attention people gave to observing the robot—was especially relevant for longer behaviors. For example, Adam on the 6th Wednesday mentioned that he noticed (apparently for the first time) that the robot visited two people in a row without returning to its home position in between. The robot had actually had this capability for four weeks by that time, but to observe it Adam had to be paying at least periodic attention to the robot for a relatively long period of time in addition to being there with at least one other person whom it had not yet visited. Another example is the robot's capability in Conditions #2 and #3 to recognize you as an individual—to get evidence of this, you have to notice that the robot visits you once but then wait long enough to see that it does not visit you again. These sorts of behaviors were less likely to be observed (fully) by our interviewees compared to shorter, simpler behaviors.

Due to these factors each interviewee only got a sample of the full set of possible

behaviors for the current condition on a given day. It was therefore difficult to compare their mental model development processes to each other directly. Instead, we treated the analysis for each interviewee as an independent case study, and then also did a looser, qualitative comparison across all six of these to put together the findings presented in the next section.

6.7.2.4 Interviewees varied widely in motivation to understand the robot

Adam clearly had the highest motivation of the six interviewees to figure out how the robot worked and what was the purpose of the study. This started with the in-class announcement about the study before he had even seen the robot. By the interview on the 4th Wednesday he said, "every day I try to figure out what's going on with it", and later referred to the "entertainment value" of experiencing all the behavior changes—"it's like a puzzle and you wanna figure it out." This extreme motivation probably increased the amount of attention that Adam paid to the robot, how much he thought about its behaviors, and how often he reconsidered or revised his mental model of it.

Some of the other interviewees avoided or even ignored the mobile shoe rack at the beginning of the study. In particular, Courtney and Donna did not put their shoes on the robot for the first three times they saw it, and Elsie and Frankie for the first two times (see Table 6.2). Donna did not even enter the taped-off area during those first three class times. During her first interview on the 3rd

Wednesday she expressed a desire to observe the robot more before answering some of the interviewer's questions and said she would "be more observant" the next time, perhaps explaining her increased attention and participation from that point forward. In fact, all six interviewees were at least *using* the mobile shoe rack regularly by the end of the study. We believe this was at least partly due to feeling obligated or incentivized to have something to say for at least 5 or 10 minutes at the next interview.

6.8 Key Findings from the Cross-Case Analysis

The interviews for each interviewee were organized into a chronological account of everything they said about mental model formation while they were participating in the study. We then analyzed these six case studies for recurring themes and key phenomena that warrant further attention and study. We have organized these findings into eight categories and present them in this section.

6.8.1 Variability of How Long it took to Form a Mental Model

It seems unlikely that the amount of time it took for our interviewees to form their mental models of the mobile shoe rack will generalize very far outside the particular setting and robot use case of this study. We did observe, however, considerable variance between our six interviewees. Some people noticed and had developed explanations for new robot behaviors after one or two days of experiencing a new

condition. For example, Adam and Bill understood by the end of the 1st Monday that the robot stops in front of each person and waits "for a second" (Bill's words) before moving on. Other people took more time to fully form their model—on the same day, Courtney said it stopped seemingly "at random" and didn't figure out that it was stopping for *people* until the following Wednesday. Still others like Elsie were still missing big parts of the robot's behavior in their final model at the end of the study.

Several factors could account for this variance. For example, we have already highlighted differences between the interviewees in motivation to figure out the robot in Section 6.7.2.4. Adam seemed to have the highest motivation, followed by Bill, and Courtney seemed to have the lowest (although we also noted that those with low motivation generally started participating more after their first few interviews). This motivation might make mental model formation faster and more accurate by increasing attention, thoughts about the robot, and how often the mental model is reconsidered or revised.

Interviewees' experiences of the robot also varied considerably (see Section 6.7.2.3), and could therefore be another factor accounting for how long it took to form mental models of the robot. Differences in attendance and participation are presented in Table 6.2. Two participants—Donna and Elsie—only attended class on Wednesdays. Bill the instructor left partway through the study and was replaced by a substitute, Frankie—both missed entire conditions. Also, both instructors often arrived so early that they did not see anyone else interact with the robot before class. Some of the interviewees reported ignoring the robot as they hurriedly

walked to and from class on some days, whereas others sat on the bench for several minutes watching it. As a result of missing class, ignoring the robot, or just bad luck, some interviewees did not experience certain robot behaviors until the condition had been active for several days. For example, it seems from the videos that Frankie's only opportunity to hear the "sad sound" was on the 5th Wednesday after class—she attended class 3 times when that sound was part of the robot's behaviors, but the occasion never arose for the robot to play it while she was there. It seems reasonable to assume that people form mental models of the robot based on which behaviors they actually *experience*, which in our study was not always everything in the behavioral protocol. Longer, more complex behaviors are probably more vulnerable to being missed because they require a person to observe them for a longer time period in order to understand what's happening (unless the person is able to guess the rest of an observation after observing just a part). Interaction designers need to understand this when they are thinking about how to teach their users about these behaviors—a longer demonstration would be needed, or else a verbal or pictorial description summarizing it.

Other factors that could account for variance in mental model formation time could include: motivation, memory, ability to reason about evidence, attention to detail, or other individual differences.

6.8.2 Reasoning about Evidence and Hypothetical Situations

When our interviewees could not answer an interview question with certainty they often talked through their thought process so that they had something intelligent to say. Adam reported in his final interview that the interviews helped him "reflect back on [his understanding of the robot] and ideally make better hypotheses." For some questions it seemed like our interviewees didn't reason about the evidence or draw any conclusions before this point. Perhaps they had the observations that would eventually point them to the model they report during the interview, but hadn't consciously put it together yet. Future work should investigate what prompts a person to reason about evidence and form a conclusion instead of just keeping the evidence in memory, unused.

Whether it happened during the interview or at an earlier time, we observed several different ways that the interviewees reasoned about evidence—either to draw a conclusion or to decide the evidence is inconclusive—and reasoned hypothetically about events that didn't happen but might suggest different (or clearer) conclusions.

6.8.2.1 Drawing conclusions

The interviewees made several different types of inferences—some correct, some incorrect—from the robot behaviors they saw.

Simple inference: A popular example: people would often notice how many shoes were on it and make inferences about how many people used it.

Correlation: some rather impressive connections were made between the robot's behaviors and the things that people thought were causing them. For example, Courtney on the 5th Monday reasoned that she couldn't hear the beeps from inside class because there was only one person on the bench when she emerged, whereas in the past there were more, and she could hear it.

False correlation: On the 6th Wednesday before class Frankie walked out of the yoga classroom and stood barely inside the zone watching the robot, "Just to see if the robot was maybe on a timer." After a few seconds of waiting, the time came for the robot to start driving, and it came up to Frankie. She said "it noticed me", falsely believing that her presence had triggered it.

Offering explanations for unexpected behaviors: Early in a new condition and especially at the very beginning of the study the interviewees didn't know what to expect, and were trying to find explanations for the robot's behaviors. For example, on the 4th Wednesday when the robot moved away from Courtney before she could put her shoes on it she suggests as an explanation the fact that there was "someone else in the box"—maybe it was trying to visit them. Why did she choose that explanation? Perhaps because it fit with a belief she was holding about how the robot visits people. Also, that other person's presence may have been more salient to her than other details like the shoes on the floor.

We find it interesting that almost nobody talked as if the robot makes mistakes—i.e., via sensor malfunctions or computation errors. A rare exception was Bill's frightened thought that the robot may have "malfunctioned" when it first drove directly towards him on the 3rd Monday. Perhaps existing theories from organi-

zational studies about *sensemaking* could be applied to these sorts of bewildered responses [253]. A few of the interviewees did talk about *randomness* in the robot's behaviors—once by Courtney on the 1st Wednesday when she had no better explanation the robot's movements and again by Adam on the 5th Wednesday to describe how its behaviors (in Condition #3) were less predictable and more "humanlike."

Eliminating explanations: When Adam noticed the new behaviors on the 3rd Monday he quickly eliminated the possibility that they were triggered by the fact that he was the only person in the area: "the situation didn't seem unique, I've come in late before ... so it wasn't like, just because nobody else was there that's what it was doing."

6.8.2.2 Deciding *not* to draw a conclusion

There were also several reasons that the interviewees gave for *not* drawing any strong conclusions from an observation.

Reasoning that the evidence may not be diagnostic. Donna in her final interview mentioned one piece of evidence that was not helpful to her. When asked whether the robot responded when she said "muchas gracias" to it, she said, "No, it ... it paused a tiny bit longer, but I don't know if it would have paused had I not said that." I.e., the slightly longer pause did not help her decide whether the robot responds to speech because it might have happened without it.

Considering the fact that their memory and attention are limited:

Several times somebody noticed a behavior for the first time, but reasoned that it might not be a *new* behavior—it could have been happening the whole time but they simply did not notice it.

For example, Courtney on the 4th Wednesday admitted that she was not paying as much attention at the beginning of the study as she is now, and is unsure whether behaviors like "the two different beeps" (i.e., the addition of the sad sound in Condition #2) are actually "a new thing" or if she didn't notice them before. Later in the interview she even doubts herself that there really are two different beeps.

Also, when Donna finally notices the orange circles on the corners of the shoe rack on the 5th Wednesday she considers the possibility that they were actually there the whole time (which they were): "... I didn't remember that it had little orange corners prior, so I don't know, maybe I'm just getting more observant."

Waiting for more evidence: Donna on the 3rd Wednesday chooses not to answer several of the interview questions because "I haven't had enough interaction" with it and "I will want to have some more time to engage what it does and so forth" before answering. She even shows **an understanding of statistical significance** during her final interview: when asked if the robot can recognize specific people, Donna reasons that one way to tell would be to see whether the robot always turns so her pair of shoes is easy for her to grab. Since the robot only has two sides, though, this would happen 50% of the time even if the robot were guessing randomly. She therefore reasons that, although the robot did present her the correct side today it may have been "coincidental" and that she would need

more interactions to test this properly. It is worth noting here that Donna is a retired university faculty member.

6.8.2.3 Forming Hypotheses and Hypotheticals

When interviewees found the robot's actual behaviors ambiguous they often brainstormed hypothetical situations that would be more helpful for answering the interviewer's question. These were essentially experiments—scenarios in which the robot's response would help them decide whether, e.g., the robot has a certain capability.

Reasoning about what it would do in an edge case: Donna in her final interview: "Well, so far as I know it's got an ongoing energy source. It's not like if the lights go out..." ...[that the robot will turn off.]

In the same interview Donna suggests testing the robot's idea of the boundary between person and object by introducing "a service animal" to the situation.

Another example was in Bill's final interview. He was unsure about one of the questions: "I never knew if there was something else there that was an object if it would treat it the same as a person in terms of stopping in front of it." He proposes putting a mannequin ("dummy") in front of the robot to test whether it can "know that it's not a breathing, living thing."

Counterfactual evidence. Some of these hypotheticals were counterfactuals—i.e., things that did *not* happen. The logic for these was of the form, if [this] had happened instead, then I would have concluded [that].

For example, Bill on the 2nd Wednesday says, "if it knew who I was through face recognition or something, it'd say oh! [These are] his shoes, so I'll bring him his shoes."'

Reasoning about what would be observable in different situations. E.g., Donna on the 5th Wednesday notices the painted shoeprints on the shoe rack's shelves: "... today I was able to see that there are actually some footprints ..., and I didn't observe that before because it was fully loaded."

It's noteworthy how much reasoning people do and how sophisticated some of it is compared to all the robot behaviors they didn't notice or forgot. Much of that reasoning may not have happened if not for the interviews. Inasmuch as people do rely on their rationality to compensate for sparse and low-quality evidence, however, their mental model formation becomes vulnerable to well-documented heuristics and biases like confirmation bias and self-report biases like the one reported by Nisbett and Wilson [171]. In general, the way people assign causes to events they observe (such as robot behaviors) is described by theories of attribution (e.g., by Jones [116], Ch. 3).

6.8.3 Comparing the Robot to Humans, Animals, and Other Devices

6.8.3.1 Comparisons and Associations

Interviewees mentioned other things a lot when talking about the robot. Most of these mentions were simply comparisons or associations. The most common comparison—made by Adam, Bill, Courtney, and Elsie—was to the Roomba robot. Early on, Adam compares it to a pet: "you feel like you're almost calling it like you're calling a dog or something, like, come here shoe rack, come on!"' Frankie compared the blabber sounds to "songbird noises", and Adam called them "R2-D2 type sounds." Frankie even made a rather abstract comparison: "I want more of a Pee Wee's Playhouse vibe out of it, and I'm more just getting [baggage carousel] at the airport with bags on it." Some of these comparisons may have *only* been comparisons—i.e., metaphors used to describe something about the mobile shoe rack.

6.8.3.2 Borrowing (pieces of) other Mental Models

More significantly, though, it appears that the interviewees sometimes borrowed from their mental models of these other entities to help them build a mental model of the robot. Borrowing from other mental models could influence the associations people make between different attributes and therefore the assumptions they make and questions they ask about a new system.

It can be difficult to tell when this borrowing is happening in addition to a simple comparison. For example, Elsie on the 5th Wednesday makes a comparison—"It reminded me of some of the sounds that my toy R2-D2 makes"—but then says, "it was kind of a fun little droid." This might have been just a metaphor, but she may have also borrowed something from her model of a "droid" as she was building a model of the mobile shoe rack.

Sometimes borrowing is more obvious. For example, Adam said on the very first day that the shoe rack appears to be on a Roomba. Here he believes that the Roomba is actually a component of the system he is trying to understand, so he might have guessed that his model for what the Roomba can do would be a very good model for what the base of the mobile shoe rack can do.

Later, Adam demonstrates some inference or prediction that resulted from using part of another mental model. He predicts how much weight the mobile shoe rack can move by reasoning, "the bottom part looks very similar to Chairbot, so if Chairbot can move 30lbs. . . ." He later specifies some differences between one of Dr. Heather Knight's Chairbots and the mobile shoe rack: ". . . the Chairbots look the bottom of them looks similar to that thing, I don't know if they're the same, but those were all being controlled by remote control obviously, I don't [think] that there's somebody remote controlling this, I would be very surprised if that was the case." He will either have to borrow just part of the other mental model or else modify it to account for this difference.

Borrowing Structure vs. Content. It is important to specify what about a mental model is being borrowed: just the structure, or also the contents that fill

that structure, or both. In the first example, perhaps Adam borrowed his *entire* mental model for a Roomba because he believed it to actually be a component of the mobile shoe rack. With Elsie it's less obvious—while it's possible that she also borrowed the typical attributes of a "droid", she may have just used the *structure* of the model as a framework. By a model's structure we mean the way the pieces of information are organized and connected. This could really matter when using a borrowed mental model—or a combination of several—to interpret observations and decide whether the model needs to be changed. For example, most humans who can understand speech can also speak, and vice versa—each ability probably strongly suggests the other one in most people's mental model of an adult human, and might be grouped as "language abilities." On the other hand, in an AI system those two abilities are not necessarily connected—it could have a speech synthesizer but not a speech recognizer, for example, such that it could speak about you and its environment but not understand anything you say. It could be important to know whether someone encountering such a system begins with their model of human language abilities as a template—if the system could even detect this it could predict the incorrect inferences and interpretations the person will make about its behavior and perhaps even choose different actions or give explanations to avoid costly mistakes.

In this study we have mostly focused on the *content* of people's mental models: e.g., which sensing capabilities the robot has and what its rules of behavior are. One obstacle to getting the *whole* picture is that the structure of mental models is unknown. I.e., it is unknown how they are represented in the mind—whether as

abstract properties with a range of possible values or as analogical connections to other, better-understood entities, or something else. It is also unknown whether people carry around one mental model—their favorite—or a list of possible candidates, perhaps with estimates of how likely each one is to be accurate. Answering these questions requires that methodologies for measuring the structure of mental models are identified and validated.

Future Work. Our results suggest several other key areas of further study on the topic of borrowing or reusing mental models.

First, it did not seem like the comparisons the interviewees made to the mobile shoe rack were random. They seemed to gravitate towards comparing it to a pet animal like a dog, or R2-D2. Which aspects of the mobile shoe rack (or the environment, or the scenario) caused them to choose those objects of comparison? Did a dog come to mind because the robot was subservient? ...social? ...of about the right size? Again, we aren't sure whether the interviewees who compared the robot to a dog were also borrowing anything from their mental models of dogs. To the extent that people do borrow, though, it will be important to predict which mental models (e.g., a dog, a droid) they take off the shelf.

A second question for future study is, how does mental model formation change when the robot's appearance or behavior is *designed* to encourage users to reuse the mental model from a certain entity? For example, AIBO is designed to be doglike, so perhaps users will use their mental model for a dog as at least part of their initial model for AIBO. We did not observe this sort of situation in our study

since the mobile shoe rack was intentionally designed to not resemble a human or animal or anything else that moves, at least in appearance.

Third is the question of what existing mental models people have for a robot, or for different types of robots. Indeed, all of our interviewees referred to the mobile shoe rack as "the robot" at times, which raises the question of what existing mental models they might have at least compared it to. The interviewees were familiar with robots from movies like R2-D2 as well as consumer products like the Roomba and Amazon Echo; they may have also known something about certain robotic toys (Adam has a 3-year-old daughter) or robots featured in news stories that have been used to assemble cars or perform surgeries. How are all these different devices represented in their minds? Is there a mental model for all robots, or are they separated into (perhaps overlapping) subclasses? There is a wide variety to account for in such a system of models—note the differences between remote-controlled cars, dolls that can hold a conversation, assembly line robots that do repetitive tasks, cruise control and lane-keeping systems in cars, and AI personal assistants.

A fourth question is how these preexisting models of robots can be changed, especially since the ones held by people who do not have in-depth experience with a robotic system are probably rather inaccurate. Did people who interacted with the mobile shoe rack in this study leave with a changed mental model for "robots"? If so, was it more accurate? Or maybe their experiences prompted them to break down the overgeneralized category "robots" into more specific subclasses: social robots vs. non-social robots, autonomous robots vs. remote-controlled robots, and

so on. Maybe theories from social cognition about stereotyping processes would apply here—sometimes researchers need to "borrow", too.

6.8.3.3 Intentionally Projecting Attributes onto the Robot

When she was asked to list the robot's capabilities Frankie makes a distinction. On the one hand, she says, "when we talk about the robot and the things we want it to do, we assign [animal or human] characteristics to it." She often thought of it as a small dog and gave it different names like "Percy." She emphasizes the difference, though, between these assigned characteristics and what it really is: "a shelf on wheels that has a sensor on it." If people often hold both of these two sets of beliefs, then it seems crucial to distinguish between them in theory, measurement, and practice.

On the one hand people have their mental models of what the robot can really do. This is based on thinking about the robot as a mere mechanism—pieces of plastic and metal controlled by computers to do things according to the designs of a human software engineer.

On the other hand people might also imagine or project additional characteristics onto the robot that they know aren't really there. This might just be an extra ability (e.g., pretending your dog understands what you say), a certain role or relationship (e.g., talking to the robot as if it is your butler), or even a whole new identity (e.g., when a robotic Mickey Mouse toy becomes the real Mickey Mouse). Designers often encourage users to imagine a particular personality for the robot

by making the robot look and act in a way that suggests that personality. When users do this, it influences the way they interact with it—they act and talk *as if* these imagined characteristics are real.

Again, Frankie is a good example of this: "I think of it kind of like one of my small dogs, like, "hey, guy!"" Also, even though she knows that this "animatizing" is "just in my mind", she appears to still use her schema for a dog as a template for her projected mental model of the robot: "So I guess I think of this shoe rack like my dogs..." (i.e., she uses her dogs as a template for the shoe rack) "..., like you tell 'em you're there, and then you go sit down and wait for them to come to you, essentially" (i.e., she is mapping her script for interacting with her dogs onto her interactions with the robot). Or, put in different words, she evaluates whether the robot looks and acts in a way that makes it easy for her to "think of it like a dog": "...it doesn't really respond to someone sitting down in the way that an animal or something small that's alive would respond to someone sitting down." So while Frankie does have a mental model of the robot's actual capabilities and behaviors, she also wants to pretend that it is a small dog. This certainly will influence her behavior around the robot, but might also influence the way she forms her non-imaginary mental model and talks about the robot to experimenters.

Perhaps HRI researchers should begin trying to distinguish between these two parts of a user's mental model—sincere beliefs about how the robot *really* works and projections that people know aren't real. It does seem like people know when they are pretending, at least when it is brought to their attention (they may have been pretending without realizing it). Presumably you could ask a person about

each mental model separately: the real life one without any pretending, and the one with imaginary stuff added on. People know the difference between a question about an actor wearing a Mickey Mouse costume and a question about Mickey Mouse the character.

6.8.4 Attributing Sensing Capabilities without Visible Sensors

Bill, Donna, and Elsie came to believe that the robot could sense distance or motion even though they had not noticed any visible sensors on the robot. Perhaps they inferred this from the robot's behaviors, but if they had seen the sensor and recognized it as a video camera perhaps they would have attributed more vision capabilities to it. This suggests the importance of "sensor transparency" [209] in robot design—among other things, so that people can ask well-informed questions about their privacy before consenting to interact with a robot.

On the other hand, Adam, Courtney, and Frankie all assumed that the robot could not hear. Why not? Perhaps they could account for all the behaviors we gave the robot using only vision or motion sensing, and were content with that simpler, more parsimonious model. Perhaps it was because the robot doesn't resemble a human or animal, nor did it have anything resembling ears. Future work should work to discover why people attribute some sensing capabilities but not others to a robot, even without obvious cues to help them. In cases wherein people are unlikely to make the right attribution, robots could be designed to help. For example, the mobile shoe rack did not use the webcam's microphone for any of its behaviors,

but it was still recording; if it had a display indicating sound levels that might remind people about the microphone.

6.8.5 Judging whether the Robot is Autonomous or Teleoperated

An important aspect of one's mental model of a robot is whether it is autonomous or teleoperated. Participants were divided over this in the robotic trash barrel study [162]: of the 28 people who commented on whether it was autonomous, 16 thought it was (often commenting that it was "clumsy") and the other 12 thought it was not (often describing its actions as more calculated or purposeful).

We tried to make our robot seem autonomous, but worried that people might suspect that there was a man behind the curtain. On the contrary, all six of the interviewees were surprised to hear during the debriefing that the robot was being operated by a human. Donna was the only person who ever suspected anything: on the 5th Wednesday she mentions that she knows "there is both visual and auditory that's being captured", and says it's possible that "somebody is actually remotely doing something" even though her best guess is that there is "some kind of sensing device on the robot itself."

The other five interviewees never suspected that there was a human operator. Some of them even got hints that could have caused them to suspect it.

Adam was the best example. He said he "never expected" it, but was also the most suspicious interviewee. He said that every time he came to class he was trying to figure out what we had changed about the robot. He actually thought

he might have become "hypersensitive" to small changes in the robot's behaviors: "am I actually picking up on things that are happening, or is my mind just doing these things?" He even develops what he calls "conspiracy theories" by the end of the study. For example, on the 6th Wednesday he noticed that the robot did not visit a certain young man sitting on the bench. He began to suspect that he had been "planted there" by the study team, and then became "more and more" suspicious of this as he noticed certain people who were "kind of out here consistently." He even admits that his experience with one of Dr. Heather Knight's Chairbots on campus should have made him suspicious because the "the bottom of the Chairbot looks just like the bottom of this robot", he "saw people controlling them remotely", and he even asked one of them, "can it move autonomously?" and got the response "no, not yet." He admits he should have concluded that, "if they didn't do it with these chairs [i.e., the Chairbots] how could they do it with the shoe rack?"

Courtney also saw a Chairbot on campus and thought, "'oh, there's a person driving,' but I assumed this one wasn't that way? Possibly because I didn't see anyone [driving the shoe rack]." She also on the 3rd Monday suggests adding a remote control function, apparently without suspecting that it already has one.

When the interviewer mentioned that the wizard sometimes drove closer to the bench than normal, Bill said, "I noticed that", but also that he did not become suspicious that the robot was being remote controlled.

6.8.5.1 Why didn't anybody suspect a teleoperator?

Adam thought that nobody guessed the robot was being teleoperated because "people have this idea of what a robot can do and can't do, and the functions that you picked were well within the realm of what a robot *could* do." "If you had it do some sort of crazy functions [like] recognizing you and [saying your] name or something like that I'd be like, 'ehh. That's not real.' Or, 'that's less likely to be real. . . '." Frankie makes a related point in her final interview—that her mental model was limited by her understanding of technology: "I think the abilities I assume the robot has are within my limits of what I could do with a robot I could create."

Our interviewees mentioned their existing knowledge about technology when talking about the robot. Adam and Courtney, for example, showed that they understand that computers can learn to classify things into different categories based on many training examples. Bill understood that the robot has "got a clock" that is probably "synced someplace." Adam even draws an analogy about the robot using another technology he knows about: "Yeah, so you never know, you could have a shoe rack like this that has a very innocuous job holding on to people's shoes but it could be keeping track of many other things. It's like license plate readers in intersections: nobody knows that they've gone through and their license plate's been scanned, but these things can scan tons of license plates at a time, and so you just never know that it's happening."

Users' understanding of what is easy, hard, and impossible with modern technology—

whether they are right or wrong—could impact their judgments about whether a robot is autonomous or teleoperated. We hypothesize that robots will be suspected of being teleoperated when they do things that people think are impossible to do autonomously.

Adam also described the robot's behavior as "realistic": "I think the way you had it in a very automated way that it went seemed very robotic to me." We intentionally designed the robot's behaviors to be robotic, especially at first: it drove in straight lines according to a simple, repeating pattern and typically only did one thing at a time. It will be important to understand each user's schema of how robots act—i.e., of what robotlike behavior looks like—and develop ways to control how robotic a robot acts, especially if it has a large influence on whether people believe it is autonomous.

A third possible explanation is that people didn't suspect a teleoperator because they didn't see one, nor did they see a wire or antenna that obviously hinted at one. This could pertain to the way we perceive social entities: perhaps we are inclined to attribute independent autonomy to bodies that are self-contained and not connected to another social actor. By this theory, a remote-controlled car would be easier to consider autonomous than a prosthetic arm would, and a robot connected to an operator's computer by a thin wire would present a border case. Wireless, remote control would cause dissonance between our unconscious impressions and our conscious understanding. Further study to test this hypothesis could help HRI researchers understand how people estimate the degree of connection between a robot and the people who might be operating or supervising it.

6.8.6 Experimenting with the Robot

In general, the six interviewees did not actively experiment with the robot very much; they mostly just watched it. When they were asked a question in an interview for which they didn't have an answer, they were quick to make guesses based on reasoning from what they had observed passively and from their mental models. They rarely took actions that were just to test the robot. Adam, for example, notes that although he had "opportunities", "I never tried to test out the limits of the robot, it just never really occurred to me. . . ." In fact, Adam on the 4th Wednesday says: "I've probably never actually touched it [i.e., the body of the robot], I just touch my shoes, right?" He contrasts this with his 3-year-old daughter, who he says would be "exploring it with her hands"—"she'd definitely want to sit on it", "She would probably try to pick it up", she would "grab" and "pull on" the sensor. This exposes a limitation in our sample of interviewees: none of them were as comfortable experimenting with the robot as a young child might be. Future work could focus specifically on how young children experiment with novel robots.

When the interviewees did experiment with the robot we noticed some connections between their experimentation and their mental models. For example, Adam on the 2nd Wednesday came to class late, after the robot had parked. He "tried to wave in front of where the shoe rack was" "to see if it was moving" "but it didn't move." He then interprets the results of his experiment based on his mental model: "maybe I didn't hit the trigger or whatever", which also shows

that his intention—to "hit the trigger"—was driven by his understanding that the robot *has* a "trigger" to hit. Hence, your mental model can inform how you create experiments as well as how you interpret the results. Presumably, someone also might design an experiment to help them decide between multiple possible models that they have brainstormed. If the result cannot easily be explained by any of the possible models, that person might brainstorm one or more new models that do explain it. Choosing *not* to experiment can also be motivated by a mental model. Frankie in her final interview, for example, said she assumed it can't hear, "so I didn't say anything to it."

6.9 Limitations

Participants might pay less attention and do less thinking about the robot in a purely observational study. Several of our interviewees admitted to changing their behavior because they knew they would be interviewed about it. If an interviewee came to their first interview and didn't have much to say about the robot, this often motivated them to spend more time watching the robot or interacting with it. For example, Courtney on the 4th Wednesday says, "It's kind of just been a progression in paying attention to it more, because I know we're doing these interviews so I pay attention to it more, I guess, instead of in passing." Elsie on the 5th Wednesday says she put her shoes on the robot when she might not have otherwise: "... I noticed the shoe rack and it was pretty full, but I decided I was going to use it anyway because I knew we were going to be talking about it,

. . . .” So at least some of our interviewees would have probably behaved differently, especially later in the study, if there were no interviews motivating them to make clear, detailed observations of the robot.

Conversations between people about the robot could be much more important in other scenarios. People didn't interact with each other much during our study. Very few of the yoga students knew each other, and most people kept to themselves while in the hallway before and after class. This means that there weren't many opportunities for people to share observations or beliefs about the robot. If Donna had mentioned the possibility of someone remotely influencing the robot's actions to Adam, for example, he might have become much more suspicious of a remote operator than she did. It could be important to understand how these and other phenomena like rumors (which might change as they spread) and a group-constructed persona for the robot (e.g., assigning it a name and personality) could influence mental model formation.

Introducing people to a new robot could include more initial explanation or training. Our participants were told almost nothing by the study team about how the mobile shoe rack looks and operates. Adam, at least, was aware of this when asked at the end of the study about how he formed his mental models of the robot: “. . . so all these [hypotheses about how the robot works] have come from just the observations and my interactions with it. . . .” We did not want to give participants too much help forming their mental models, but in reality people often see television ads, YouTube videos, or news stories about robots before they first interact with them. They might also read written descriptions, see still

photographs, or hear about the robot from a friend or colleague. This initial information about the robot could be received as authoritative (e.g., if it's from the robot's manufacturer) or not so much. It could be important to understand how this information sets up initial expectations of the robot[1] and how people reconcile observations with rumors when the two don't match.

Limitations in demographics of interviewees. The interviewees who volunteered for this study were relatively well-educated people, several of whom worked in academia. Perhaps they were therefore more likely to hypothesize explanations for the robot's behavior and to evaluate these explanations in light of their observations. They might also have been more likely to avoid being seen treating the robot as a person or animal if other people might see that as childish or ignorant of the fact that robots are not *really* alive. Our interviewees were also relatively open to new technologies according to their responses to our questionnaire items.

The choice of a university campus in a small town with low crime rates might have influenced how much people paid attention to the robot and noticed details about it. This hypothesis was suggested by Frankie, who was surprised by how many people did not "look up" at the robot or talk about it. It seems she expected them to be more alarmed by a "weird shelf moving around" and to at least ask somebody else what it was. She theorizes that people in this area "just don't notice their surroundings" compared to people from her home town, where "their surroundings aren't always perceived as safe." The extent to which people are

[1]Prior research on the effects of these expectations includes the work by Paepcke and Takayama [177].

suspicious of the robot (or its operators) or are worried about their safety or privacy around it could influence the way they form a mental model of what it is and how it works.

6.10 Additional Suggestions for Future Research

The previous two sections, which present our findings and the limitations of our study, contain many ideas for future research that arose—perhaps for the first time—from this study. This section presents several more suggestions that have not already been mentioned.

6.10.1 Additional Research Questions

What about nondeterministic behaviors? Although the wizard sometimes made mistakes, the robot's behaviors were supposed to be completely deterministic and predictable. It would be interesting to study how people reason about robots with more stochasticity (randomness) in their behaviors, or that change their rules of behavior more gradually over time (e.g., via learning). It seems from our results that people might be slow to notice that a robot's behaviors are changing, especially if they only observe it for short periods of time. Future work could explore how people make this judgment—what cues they look for, and how they reason about it.

Is there a way to present behaviors to control mental model forma-

tion? Presumably it matters which behaviors people see and which order they see them in when building a mental model of the robot. Is it possible to unveil a robot's features in a certain order to increase the accuracy of the final mental model, or help people develop mental models more quickly? For example, it might be a bad idea to show a sophisticated ability—e.g., like if your robot can speak—in the first interaction because people might infer other abilities that your robot does not have [221]. Note that this is different than teaching someone about simpler behaviors first to make the more complex behaviors easier to understand—here we are talking about what each behavior might *imply* about the robot's abilities, thereby influencing mental model formation. Future work should include multi-interaction studies in which we manipulate the order in which the behaviors in the robot's repertoire are revealed. Maybe the robot would also reveal negative behaviors, too: it could intentionally fail to do something or otherwise show what is *not* in its repertoire. Understanding the impact of these manipulations would help designers control mental model formation.

6.10.2 Lessons Learned about Study Design

If we were to change the methodology used for this experiment and run it again, we envision taking it in one of two directions:

1. We would loosen some of our experimental controls and do more active exploration. We realized when the robot ran out of battery for the 4th Monday that when things don't go according to plan it is also an opportunity to see

how our participants react to abnormal circumstances. We think it would be fruitful to actually plan intentional malfunctions or failures to see how participants handle them—real robots do malfunction, after all.

We would also consider improvising some interactions to test the limits of a particular person's mental model of the robot. For example, if they currently believe that the robot is autonomous, we could try to find behaviors that would convince them that it is teleoperated. Or, if they believe the robot cannot hear, we could try responding to sound in different ways until they believe that it can. This would be a more active, targeted approach to understanding what influences specific elements of a mental model.

2. Alternatively, we would attempt a more direct comparison between participants by controlling which behaviors they experience. One way to do this would be to keep it to one participant encountering the robot at a time. We would probably not try to control which behaviors they *notice*, but would instead measure that carefully and adjust for it in the analysis.

   We could also control some parts of the mental model formation process to make the study more focused. For example, we could try to encourage a certain template model for the robot by designing the its appearance and behaviors to evoke people's existing mental model of a particular thing, like a pet dog or a human butler. Pilot studies would be required to make sure that this desired impression is made clearly and uniformly across participants. Controlling this source of variance might help provide a clearer picture of

other parts of the mental model formation process such as updating based on interactions with the robot. On the other hand, findings from such a study might not generalize to robots that suggest different templates or that do not suggest a template at all.

We also learned how influential the design of the robot's appearance and behaviors is to how participants respond to it. In future work it will become important to produce robots that target a certain phenomena of interest by evoking a certain response in participants with high precision and consistency. For example, you might want a robot design that ambiguously suggests two different mental model templates—e.g., either a pet dog or a human butler—to study how people choose between them. Or, you might want to design two slightly different behavioral modes for the robot that are activated based on some feature of the environment, and see how long it takes people to connect the behaviors to that environmental feature. Future studies of mental model formation will require design experts on the research team so the robot and its interactions can be better aligned with the research goals in this way.

All of these recommendations are intended to make longitudinal, in-the-wild studies—which are often avoided because of how long they take and their lack of focus—more targeted and efficient.

6.11 Summary of Findings and Recommendations

We have presented findings from the first long-term, in-the-wild study of mental model formation about a novel robot. We analyzed 28 interviews of six participants during six weeks of interactions with a "mobile shoe rack" working outside a yoga classroom. We can summarize our findings in two main conclusions:

1. **Different people can experience a new robot very differently, especially over multiple interactions.** Each participant had a different set of observations: out of the robot's list of possible behaviors, only a subset were triggered while a particular participant was watching, and only a subset of these were noticed. Each participant also had a different preconception of what is technically possible for an autonomous robot. Finally, participants varied widely in motivation to figure out the robot, and took different amounts of time to form their mental models.

2. **Mental model formation is a complex process consisting of multiple components and sometimes yielding surprising results.** A mental model of a robot might start from an existing mental model that is used as a "template." Given some observations of the robot, participants then use a variety of types of reasoning to draw conclusions—we documented eight different types. Inferences can be indirect, like inferring the presence of a sensor without actually seeing it by using the robot's behavior. Participants also make predictions about the robot's behavior and design experiments to

test their mental models, although they do not always run these experiments when given the opportunity.

We documented a few of our participants' specific beliefs that warrant further study on their own. For example, we speculate about how none of our six interviewees suspected that the robot was actually teleoperated despite receiving clear hints to suggest this. We also gave a report of someone knowingly maintaining a make-believe persona for the robot alongside the real one.

All of these findings demonstrate the fruitfulness of long-term, in-the-wild studies, and we end by giving study design recommendations for making them more efficient and focused. Our research topic—mental model formation—is fundamental to many important areas of HRI research, such as user interface design, human-robot collaboration, trust, safety, and privacy. We hope this report will serve to bootstrap more research on how people form mental models of robots.

Acknowledgment

The study team wants to thank Dr. Sam Logan's lab for their welcome and the use of their space. We also thank Sonny Goodnature for helping us set up our equipment. Thanks also to "Bill" and "Frankie" for opening their classroom, and to all six interviewees—including "Adam", "Courtney", "Donna", and "Elsie"—for their time, openness, and thoughtfulness. The author wants to thank his colleagues

for helping to design this study: Duy Nguyen, Sabrina Bradshaw, Alison Shutterly, Wendy Xu, and Marlena Fraune.

6.12 Lessons Learned

This work has confirmed to us the importance of studying privacy phenomena in HRI over multiple interactions to get past novelty effects and in natural settings to enhance experimental realism. We recommend that this be part of the standard of rigor for privacy-sensitive robotics research.

When we designed this study we were interested in issues of sensor transparency (i.e., how obvious the robot's sensing capabilities are to users) as the main point of relevance to privacy concerns. We reported some findings about this in Section 6.8.4. Some people attributed motion or distance sensing capabilities to the robot without actually seeing the sensor (a webcam), whereas other people assumed there was a *lack* of hearing ability without doing a comprehensive search for a microphone. It appears the robot's behaviors (or perhaps also its appearance) were enough to imply a likely set of capabilities to users.

Perhaps more striking is the fact that none of our six interviewees suspected that the robot was teleoperated. It seems it is relatively easy to watch people over a live video feed without causing suspicion. Note also that the robot was very close to these people, many of whom were taking off outer layers of clothing or emerging, sweaty, from an hour of exercise. Despite this, none of our interviewees reported feeling uncomfortable when the robot was pointed in their direction, perhaps be-

cause they did not believe there was a human watching them remotely! It seems crucial to understand in what situations people will make themselves vulnerable to privacy violations by assuming that nobody is monitoring a robot's sensor feeds. We suspect that the robot's consistent behaviors and lack of wires leading around the corner, along with the absence of any study team members during operation, may have been important factors.

7 The Future of Privacy-Sensitive Robotics Research

In addition to defining the emerging research area of Privacy-Sensitive Robotics (Section 2.4), performing an initial literature review (Chapters 1 and 3), and reporting findings from a series of empirical studies (Chapters 4, 5, and 6), we now provide some vision for the future of Privacy-Sensitive Robotics research. Planning for the future of this new area of research is non-trivial due to its unique nature, especially its multidisciplinarity. For one thing, it involves aligning engineering technology with a social value that's hard to define and operationalize, bringing together two realms that are usually separate. In addition, Privacy-Sensitive Robotics touches many different application areas, and we will need people from many different disciplines to accomplish our goals in all those areas.

This final contribution has three parts. First, we present a roadmap for Privacy-Sensitive Robotics research—a set of both basic and applied research directions that we recommend as next steps. The second part came out of the first workshop on Privacy-Sensitive Robotics, which we organized at HRI 2017. We present seven themes chosen to cover the whole breadth of Privacy-Sensitive Robotics research by workshop participants from a variety of disciplines. Third, we list application areas that would benefit from robots that are more privacy-sensitive and the types of expertise that will be needed to build them. This last part will end by emphasizing

the importance of collaboration—including with many *new* collaborators who have not yet been involved in HRI research.

7.1 A Roadmap for Privacy-Sensitive Robotics

In this section, we propose a roadmap for privacy-sensitive robotics, setting out what we see as the most important research questions that need to be addressed. Our recommendations have been adapted from work that was presented by Rueben and Smart [195] at We Robot 2016. We begin with *basic research* that will be foundational for any privacy-sensitive robotics application, then list some *applied research* ideas that begin to approach real-world solutions for promoting privacy in human-robot interactions.

7.1.1 Basic Research

7.1.1.1 What is Privacy? A Privacy Taxonomy for HRI.

The first question that we need to answer for privacy-sensitive robotics to move forward, before we even begin to think about robots or the technology involved, is what we mean by the word "privacy." As we have said in Sections 2.1.3 and 2.1.1.4, there are many different ideas that could be described as "privacy," from the common notion of having information you don't want revealed to the notions of personal space and solitude. We propose to complete and validate the preliminary taxonomy of privacy we presented in Section 2.1.3 for a human-robot

interaction context. This would give us a well-defined language for identifying the facets of privacy that are relevant in different application areas. Carefully defining each concept would help researchers to operationalize it into concrete, observable phenomena that can be measured unambiguously. Our taxonomy in Section 2.1.3 serves as a starting point for this in that it collects facets of privacy from a broad survey of the privacy literature. The next steps are to make sure our list of facets is complete, deal with facets that overlap or seem to belong in multiple parts of the taxonomy, and then validate that our taxonomy matches the way people really experience privacy.

7.1.1.2 How do People Think of Privacy? Identifying Key Factors that Mediate Perceived Privacy Violations in HRI.

Armed with a way to talk about privacy, the next step is to determine what people think about privacy in the context of robots. What are their concerns? What types of privacy violations worry them? Is it even something that they have thought about?

Indeed that is the first question: we need to look at how much value people put on privacy protection, if they even want it at all. Many people are willing to trade privacy violations for convenience—using location services on their mobile phone, for example. Understanding the (perceived) cost of privacy violations in relation to the (perceived) benefits of a privacy-violating service will further help

us understand what we need to work on, and how important it is to people. This will let us focus our finite resources where they will have the most impact.

When a person feels that his or her privacy has been violated when interacting with a robot, it is important to ask which aspects of the person, the robot, or the surrounding situation influenced that feeling. Here we are concerned with subjective or perceived privacy, not, if such a thing exists, objective privacy. There are probably multiple factors that influence this perception for each facet of privacy; for example, Takayama and Pantofaru [234] give experimental evidence for several factors that impact personal space, including experience with pets and robots, robot gaze direction, and the sex of the subject. It is important to identify what key factors mediate perceived privacy violations so robot designers with good intentions can avoid offending users. Beneficent designers like these would also want to know what sorts of scenarios give people a *false sense of privacy* so as to avoid them; an informed awareness about this would also help policymakers and judges make better legal decisions about robots.

Answering this question would involve first taxonomizing the idea of "privacy" as described in the previous section, then enumerating and testing the plausible mediators for each construct. Mediators could be part of the human, such as personality traits, experiences, or demographics; the robot, such as morphology, behavior, or appearance; or the surrounding context, such as the task at hand, the way the robot is introduced or framed, or environmental factors such as ambient noise or temperature. The goal is to identify the controlling mediators for each construct under the "privacy" umbrella. We want to be able to accurately predict

which types of privacy will be of concern for a given HRI scenario, or, given that we are concerned about a certain type of privacy violation, to minimize the risk of it occurring.

7.1.1.3 How do we Evaluate Privacy? Developing Standard Scenarios and Measures for Privacy-Sensitive Robotics.

Existing work in privacy-sensitive robotics (see Section 3.3) has not included much collaboration. In particular, each of the user studies conducted thus far has featured a unique experimental design, custom-made for the specific purposes of that one study. We know of no *replications* of privacy-sensitive robotics studies. The standard taxonomy of privacy terms discussed in Section 2.1.3 is crucial for ensuring that researchers are talking about the same things, but two further standardizations—of *scenarios* and of *measures* of privacy—will also help promote the collaboration between researchers that is so essential.

By *scenario* we mean both the setting and the interaction or task being studied. This includes the briefing given to participants, the type of space (i.e., public or private, large or small, home or office or outdoors), the surrounding sights, sounds, and smells, and of course the type of robot being used and its behaviors. There are two purposes for developing standard *scenarios* to be reused between studies. First, in basic research, we want to perform valid replications to confirm findings. We also want to check whether results *generalize* to different types of people (part of *external validity*), which requires holding the scenario constant. Second, in

applied research, we want to make valid comparisons between privacy protection systems. To say that system A is better than system B requires that we test them in comparable scenarios.

A standard privacy *scenario* could be any human-robot interaction in which privacy is a concern. Of course, it should be meaningful to real-world applications and reasonably easy to replicate. An example scenario would be for a janitorial robot to clean a bathroom that might have people in it. We could define a series of tasks to be completed in order (e.g., clean the toilets, polish the mirrors, take out the trash) and one or more configurations of human occupants (e.g., sitting in a stall, washing one's hands, walking in the door). A given scenario could be implemented in a natural setting, a laboratory setting, or in a computer simulation, perhaps with humans participating via virtual reality[1]. These scenarios can used both as a series of challenges to motivate work in privacy-sensitive robotics, and as benchmarks for performance testing like toy domains are in the field of machine learning. We add as a warning, however, that privacy-sensitive robotics promises to be an especially hard area for which to find replicable scenarios. Certain kinds of privacy—territoriality and personal objects come to mind—seem to depend on the *relationships* people have with spaces and objects. Controlling these relationships between replications of an experiment or even between subjects in a single experiment seems especially difficult in this area of research.

This brings us to *measures* of privacy-sensitivity in robots, which we hold to be

[1]This may be the first practical use case that has been suggested for using the restroom in virtual reality.

equivalent to privacy upheld and violations avoided for humans. These measures will be common, reusable operationalizations of the privacy constructs defined by the privacy taxonomy that we recommended above. It is important to establish and validate measures that can be reused by multiple researchers for replication or convenience. *Objective measures* will be concerned with the robot's actions and with the situation as it physically occurs, whereas *subjective measures* will be concerned with how real people think and feel. Both types are useful, but when constructing the latter we should first review the experimental social science literature (e.g., Nisbett and Wilson [171]) for best practices and additional techniques that are not yet well-established in HRI research. For example, long-term studies might involve ethnographic observation and experimental settings that are less controlled than we are currently comfortable with as a field.

We will also want to pay attention to the distinction made by [79] between the "watcher" and the "watched," as well as a new, third role, the robot *operator*, who is probably a "watcher" but also actively controls the robot[2]. People in each of these three roles could experience privacy differently in a given human-robot interaction, so good measures will distinguish between them. It will also be important to control the users' level of understanding of the robot's abilities and lifelikeness. This is one area in which scenarios and measures converge, since the briefing given to participants will affect what is being measured, and with what validity. Uncontrolled framing and poorly-worded questions can easily skew

[2]We owe the observation that a single individual could hold several of these three roles to Ross Sowell.

the results, especially for experiment designers not intimately familiar with the measurement instruments.

7.1.1.4 What are our Tools? Implementing Privacy Protection on Real Robots.

Once we understand people's privacy concerns and how they evolve, we can start to map these concerns onto technology that protects their privacy. If people are, for example, concerned with their image being seen by a remote robot operator, we can draw on the field of computer graphics to provide visual filters (redaction, blurring, or other more sophisticated techniques; see Section 3.2.1.1) to obscure the sensitive parts of the image while retaining enough information for the remote operator to complete their task. Similar techniques from other fields could be used to address other types of privacy concerns.

These techniques will have to be implemented and combined into robotic systems with usable user interfaces. Systems integration could introduce new problems due to limited computation or memory, conflicts between different privacy protection techniques, and security vulnerabilities. These systems will then need to be tested to verify that they work as expected. Long-term experiments in natural settings, though time-consuming, will be important if we are to trust these systems.

7.1.1.5 Using a more Mature Field as a Guide.

When figuring out how privacy-sensitive robotics should grow as a new field of research, it might be helpful to look for similar fields that have successfully matured already. Perhaps the most natural comparator to consider is computer security, as it is often mentioned in the same breath as privacy is because "privacy and security" is a field of computer science research (though "privacy" here is typically just digital *information* privacy). At first, the comparison seems to fit. Computer security involves a core of technical research into algorithms and techniques as well as a specialized application of these techniques in different contexts like banking, internet browsing, and file encryption. Similarly, there are core technical elements to privacy-sensitive robotics, such as the algorithmic blurring of faces in an image stream. These elements are then applied in a particular context, such as a telepresence system, as appropriate. Not all elements are appropriate for every context, just as with computer security. We believe, however, that this metaphor is ultimately misleading because security is not nearly as *personalized* as privacy is. In computer security, relatively universal, standard solutions can often be sufficient. Most browsers use 128-bit encryption, for example, even if the data are not particularly sensitive. When thinking about privacy, however, we believe that protections will need to be much more individualized, and in nuanced ways.

With this in mind, we propose *accessibility* as a model on which to base our thinking about privacy-sensitive robotics. Accessibility refers to the design of artifacts and services that can be used by people with disabilities. It is not a single

set of techniques, but a collection of designs, features, and approaches that can be combined to accommodate any particular set of disabilities. No two people with disabilities are exactly alike, just as no two people have the same set of privacy concerns. There are, however, broad classes of disabilities, just as (we expect that) there are broad classes of privacy concerns. Accessibility is improved when the accommodations line up with the particular disabilities, just as (we claim) privacy protections will be more satisfying when the particular technical measures align with an individual's privacy concerns. Finally, just as with accessibility, the final decision about a certain feature's appropriateness is made by the person who is using it, not by some objective measure.

We believe that using accessibility as a model will help us think about privacy-sensitive robotics and, in particular, how it interacts with application domains. Privacy-sensitive robotics needs to be studied in its own right because, though it might look different when applied to different situations, there is core, general knowledge to be gained as well. This is why accessibility is treated as a standalone field, and privacy-sensitive robotics should be as well. On the other hand, privacy-sensitive robotics is barren until applied to a specific privacy construct and a specific context. In fact, *applying* the general knowledge and best practices in a context-aware way is an especially large part of privacy-sensitive robotics, just as it is for accessibility. A deeper dive into the history of accessibility research could help us decide how privacy-sensitive robotics should *progress* as a field, which might currently be unclear due to its unique nature.

7.1.1.6 Challenge Problem Frameworks.

Another question to consider is whether we should have large challenge problems in privacy-sensitive robotics. In other areas, challenge problems have helped guide the community, focus resources, and provide a common purpose. It is easy to imagine challenge problems being useful here for some of the technical details, such as building a faster face blurring system for image streams. Once we move past implementation details to whole system performance in real scenarios, however, things become harder. Privacy varies from person to person—does posing individual challenge questions even make sense? Does making the problem general enough to pose as a challenge also rob it of its usefulness in the real world?

Part of the problem lies in the evaluation of privacy protection. In many areas of robotics, there is an objective measure of performance by which you can compare two systems. In robot localization, you can measure how far your position estimate is from the actual position of the robot. In the DARPA Grand Challenge, you can measure how long it takes for the robotic vehicles to travel a pre-specified course. When assessing privacy protections, it is harder to have such a crisp metric that is so easy to calculate. This makes designing challenges hard.

It is our responsibility to rigorously define what we mean by privacy protection in a particular context, and to come up with ways to assess it. Depending on the case, assessment could be objective or, via the use of human judges, subjective. This highlights the need for a taxonomy of privacy terms, as we have already

discussed above in Section 7.1.1.1. Once this taxonomy is in place, we will be better-equipped to design challenge problems for privacy-sensitive robotics.

If challenge problems do prove useful in this field, they should account for the fact that privacy is inherently an adversarial notion. One person is trying to protect their privacy, while another is trying to violate it, although perhaps not intentionally. This naturally leads us to think of privacy challenges as being two-sided: one side trying to protect privacy in some limited, well-specified context and the other trying to violate it by overcoming the protections. In addition to making for a more exciting challenge, we believe that this sort of challenge would better test our technology by exposing its weaknesses.

7.1.1.7 Re-thinking Intentionally Anthropomorphic Robots.

We often program anthropomorphic behaviors into robots as aids to human understanding and interaction. A robot might point its head at you to signal attention or scratch its head while it is processing something. Those cues are easy for humans to understand, as well as attractive and compelling. Humans like to make images of things—we are *artists*—including images of ourselves. But making robots humanlike also causes a certain class of problems due to the ways robots and humans are fundamentally different. Anthropomorphic behaviors encourage people to project human characteristics onto robots, even ones that are not the case. For example, imagine a robot vocalizing that it is planning a path with a thoughtful, "hmmmmm!" Just using its voice might make humans think it can speak, or even

understand speech. Even robots that can understand and speak some words can only do so to a very limited extent, but naïve users don't know that, and might assume that it understands and responds to unstructured speech. This is just one example of how anthropomorphic behaviors—here, speaking—can cause users to *infer* things about the robot that aren't true, thereby stumbling into a misunderstanding. Richards and Smart [191] call this mistake *The Android Fallacy*, and add that it could lead lawmakers to treat robots that look or act like humans as if they have free will or fallibility. The authors argue that we must remember that robots are machines, and ought to be held to machine standards.

The Android Fallacy also causes problems in settings where privacy matters. Anthropomorphic behaviors might trick users—perhaps intentionally—into believing that a robot's sensors have the same limitations as human sensors do, or that a robot is socially aware and privacy-sensitive when it isn't. For these reasons, privacy-sensitive robotics researchers should re-examine the practice of anthromorphization in robotics. Particularly, we are interested in the adverse effects for privacy of framing robots as humanlike (see Darling [54]). What sorts of robot appearances, behaviors, and descriptions cause the false inferences that constitute The Android Fallacy? When does this become a privacy risk? In those cases, can we think of non-anthropomorphic ways to accomplish the goals of the interaction?

This research direction can be expanded beyond the issue of anthropomorphization. In fact, we are concerned about any wrong conception of a robot formed by users. Perhaps the more general question is, what sort of robot appearance, behavior, and description encourages observers to form an accurate mental model of

how the robot works? Is it a good idea to reuse pre-formed mental models (e.g., of a human, a dog, a calculator, a hammer) as local approximations of what robots are like in certain scenarios, or do we need to foster a mental model completely unique to robots? If the latter, what might that model look like?

7.1.1.8 Can Privacy Protection Make Privacy Worse?

Many of the privacy protection methods given in Section 3.2 and in the applied research directions below have the potential to call new attention to private objects, regions, or people. If a robot blurs out a particular person or avoids a certain room, onlookers and operators might begin to wonder why that person or room is so important. If the robot avoids an object that the remote operator can't see, that person might even be able to infer its location. What's worse is, these inferences of value, presence, or location all help a malicious user to try breaking through the privacy protection. Research is required to investigate when and how this phenomenon occurs, and how to prevent it when possible.

7.1.2 Applied Research

7.1.2.1 Robot Transparency through Privacy Warning Labels.

We want privacy-sensitive robots to be *transparent* about their actions and inner workings. This means that what the robot *appears* to be doing and thinking

matches what it is *actually* doing and thinking. We also want to make this clear and obvious to human observers, with a minimum of possible interpretations.

One potential transparency mechanism is a standardized labeling system to disclose privacy risks to people that are around robots. Sometimes it's hard to tell what a robot is capable of; labels on the robot's outer casing could indicate whether the robot can record video or sound, recognize peoples' faces, or connect to the Internet. If the robot's capabilities depend on which software packages are currently active, an outward-facing screen on the robot's body could indicate whether, e.g., the robot is programmed to respect personal space as it speeds down the hallway. These labels could also be broadcast to nearby devices so that people could check their smartphones or monitors on the walls to see relevant warnings about nearby robots. Users could also interact with the labels to disable certain capabilities. Each label could light up when that particular capability is in use, e.g., when the robot is looking for faces it recognizes. This system might inspire the sort of easy dialogues we want in a privacy protection interface: users can be warned of risks, opt in or out, and even receive notice of when preferences conflict or certain robot capabilities are too important to disable.

Research will be required to develop a visual language for all the robot capabilities and possible harms relevant to privacy. Kelley et al. [123] describe their design process for a system that was inspired by nutrition labels—efforts to do this for robots should consider their example and others that are similar. Prototypes of an active privacy labeling system for robots should be tested with real users in realistic scenarios for usability, understandability, and trust that privacy prefer-

ences are being honored. At first glance, *standardization* seems like a good way to help members of a diverse public understand a broad spectrum of robots.

7.1.2.2 Robot Transparency through Graphical Interface Elements and Behaviors.

Besides an explicit labeling system, robots can also become more *transparent* through what they make apparent to humans through graphical interfaces and robot behaviors. This could increase user trust in a privacy protection system by allowing users to *see* their preferences visualized or acted out by the robot. Users could also see sensor data and data products to better grasp the robot's sensing modalities and limitations. For *visualization*, augmented and virtual reality (AR and VR) devices could help immerse users in the world as seen and processed by the robot. The robot could also project images into the environment, especially flat surfaces nearby, such as the ground around its immediate footprint. Regardless of display modality, visualizations help users confirm their privacy preferences and open up the robot to increased introspection.

Robot *behaviors* can also make privacy protection more transparent. For example, the robot could avoid private objects and regions more obviously, making it clear that the privacy restriction is changing what would be the robot's normal behavior. Here perhaps we can apply work on legible robot motion, e.g., by Dragan and Srinivasa [62] for manipulation and as reviewed by Lichtenthäler and Kirsch [150] for navigation. This might be especially important for respectful manipula-

tion, since it may be inappropriate not only to touch something, but even to reach towards it. Research is required to evaluate the application of legible motion to privacy protection, however, since legible reaching and respectful reaching, for example, might be different. Other transparency-promoting behaviors might include turning off lights on a sensor and pointing it in a useless direction to show that privacy settings are being obeyed. Research in graphical and behavioral transparency is required to find the best ways to help users understand the robot's actions enough to be confident about the privacy protection system.

7.1.2.3 Robust Visual Privacy Protection for Telepresence Robots.

Another future research direction is to make a robust visual privacy protection system for telepresence robots. The goal is to allow local users to mark people, objects, and areas to avoid, and the robot will honor those privacy settings for all remote users. Part of this research would be on the effectiveness of different interfaces for specifying privacy preferences. Rueben et al. [198] give some initial, qualitative findings, but many questions remain unanswered. Which types of interfaces are easier to use over long periods of time? Which promote trust that the privacy protection system is working? How can interfaces make privacy settings clearer, and when are there opportunities for mistakes? If physical markers are used, can the robot robustly acquire and reacquire them in all different circumstances?

This new research could also aim for the most complete privacy protection

technique: image replacement. As introduced above in Section 3.2.1.1, OctoMap and RGB-D SLAM could be used to maintain a somewhat high-fidelity 3d color map of the robot's environment. This map could be built with private persons and objects absent, and then used to provide convincing replacements when such things are present.

It is impossible to *guarantee* privacy with any real system, and *probably private* might be a good enough setting for some users. A probabilistic framework could be used to provide these soft guarantees based on user preferences. Such a framework would use the robot's certainty about which regions of its video feed are private to decide how aggressively to apply filters. If the position of a private object is modeled as a Gaussian probability distribution, a privacy filter could be centered at the mean position and inflated beyond the object's actual size to provide the user with more certainty. This would filter the entire screen whenever the robot is not localized within its map, which is a rational action given those circumstances.

User studies would be necessary to answer our research questions for the replacement filter and also the probabilistic privacy framework. First, do they work? Are they seamless, usable for various tasks, and convincing to the remote operator? How does the local user know when to *trust* the system—does this require extra feedback channels?

7.1.2.4 Respecting Personal Space.

This is more complex than keeping beyond a constant distance from each person; here we again refer to the work by Altman [10] and Burgoon [39], discussed above in Section 2.2.3.1. Personal space preferences change with the situation. Elbows can touch in a crowded hallway, but not in a deserted alleyway. Certain activities, like exercising or doing delicate work, demand extra space, whereas others, like dancing or taking in a view, invite close company. Emotional states, too, communicate the need for space or openness to proximity. There is also evidence for various sex differences in personal space preferences (see La France and Mayo [135] for a review). Territoriality probably matters here as well: the situation will change when it becomes *my* house, *my* room, *my* desk, or *my* computer. A personal space-sensitive robot should be able to perceive all these factors, reason about the situation, and move so as to respect the personal space of everybody present given the context.

How, then, would the spatial preferences be implemented when we need to plan a path to a goal? The work by Lu and Smart [151] has already been discussed in Section 3.2.2.3 for modifying cost maps to account for personal space. More work is needed to account for all the contextual factors. Since human actions are difficult to predict and the situation can change suddenly, the robot might need to replan midway through a movement; we will need an anytime planner. The robot also needs to deal with the inevitable awkward situations when personal space requirements conflict and violations occur. Perhaps gestures or utterances

familiar from human social encounters would be useful here (pending the resolution of concerns in Section 7.1.1.7), such as averting the eyes or saying "excuse me" as appropriate.

7.1.2.5 Audio Privacy: Initial Exploration.

Personal privacy can be violated just by listening. Some conversations are private, some sounds are embarrassing, and sometimes people don't want to be heard. In one sense, protecting audio privacy might seem easier than for visual privacy. Robots can probably operate without sound in more applications than without images, so the easy solution is to turn off the microphone in sensitive times or areas. When this is not possible, however, audio privacy may well prove to be more difficult than it seems to be. For instance, notice the *asymmetry*: a robot could be listening to a conversation from the next room without the conversants being able to hear or see the robot. Vision, on the other hand, is usually symmetric—i.e., if I can see you, you can see me. Asymmetry certainly poses problems for the *transparency* protocols discussed above: from the next room, the conversants would not see the "audio recording enabled" warning label illuminate, nor would they see the robot's microphone extend towards the connecting doorway. Restoring symmetry here might cause other problems: e.g., a robot that hums so loudly that you can hear it from the next room would not only interfere with its microphone, but would also be loud and annoying. Short of shouting a warning ("I can hear

you talking in there!"), we still need transparent and socially acceptable ways to enforce audio privacy protection to minimize accidental or malicious violations.

7.2 Research Themes and Future Work

The first workshop on Privacy-Sensitive Robotics was held in conjunction with the ACM/IEEE International Conference on Human-Robot Interaction (HRI) on March 6th, 2017 in Vienna, Austria. The deliverable for the workshop was to make a complete list of the various subject areas within privacy-sensitive robotics, identify pressing research questions in each area, and compile these into a document to help focus the efforts of this new community. The results of this process are presented in this section.

The workshop was organized by Matthew Rueben, Bill Smart, and Cindy Grimm from the Collaborative Robotics and Intelligent Systems (CoRIS) Institute at Oregon State University as well as by Maya Cakmak from the Computer Science and Engineering Department at the University of Washington. The day's activities included two invited speakers, a set of shorter contributed talks, and an extensive ideation session wherein workshop participants brainstormed productive research directions, both from their own prior work and in response to the work presented at the workshop. The 15 participants included human-robot interaction researchers with a variety of backgrounds as well as people from privacy and security, human-computer interaction, and law. A full schedule is available along with the con-

tributed papers on the workshop website[3]. The seven research themes presented in the next few subsections (7.2.1–7.2.7) and the proposed research directions therein were drawn from a paper by Rueben, Aroyo, Lutz, Schmölz, Van Cleynenbreugel, Corti, Agrawal, and Smart [200] that was presented at the IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO 2018).

7.2.1 Theme 1 of 7: Data Privacy

7.2.1.1 Storage and Processing

Personal information may be collected, processed, stored and shared by robots and the people who have access to their hard drives. Robots can be divided into three categories depending on how personal information is processed and stored:

- *Onboard processing*: robots such as the Roomba are able to do all their information processing and storage within their body, without the need of external components[4]. Onboard processing seems to be the best solution for privacy, but it may offer low performance due to its technological limits.

- *Local processing*: robots such as Cozmo need a local computing resource like a smartphone or PC to function, although no Internet connection is required[5].

- *External processing*: robots such as Pepper or HelloBarbie need an external

[3]https://sites.google.com/oregonstate.edu/hri-2017-privacy-workshop/program
[4]https://www.technologyreview.com/s/541326/the-roomba-now-sees-and-maps-a-home/
[5]https://support.anki.com/hc/en-us/articles/236021007-COZMO-Cozmo-Basics

resource to function[6]. Such resources are usually located in the cloud and used by the company to process the data. Transmitting the data via the Internet exposes it to additional security risks. Also, cloud servers could be located anywhere in the world, and might be owned by a third party company that provides cloud storage or processing as a service. Users should be notified about which data are transmitted and stored externally and the possible risks of doing so.

The growing field of cloud robotics raises some additional privacy concerns. One application of cloud robotics is in remote robot learning [107], in which a robot is controlled by a remotely located user to teach the robot to perform tasks that are difficult to do autonomously, such as grasping. Also, some of the robot supervising roles can be outsourced to places where more human capital is available. Both of these aspects of cloud robotics pose increased privacy and security risks. If unauthorized people were able to get control of the robot, they could cause physical damage or spy on local users. Willow Garage's Heaphy project on remote robot learning ended partly due to these privacy and security challenges[7]. Privacy and security researchers are needed to help analyze all the options discussed in this section and especially to identify any special risks from factors that are unique to robots, such as embodiment and mobility.

[6]https://toytalk.com/hellobarbie/terms/

[7]https://spectrum.ieee.org/automaton/robotics/robotics-software/the-heaphy-project

7.2.1.2 Technical Enforcement

One way to enforce privacy protection is to create levels of clearance and to allow the robot to recognize and categorize people into these levels. For example, if the robot lives with a family all the family members could have permission to access family-related information. If a guest comes to the house, the robot should recognize him or her as an outsider and should not disclose that kind of information. If new information is added to the robot, it should be stored at the strictest clearance level by default.

Some privacy concerns can be lessened or even avoided completely via technical solutions. For example, the robot could avoid certain areas, perhaps during specified hours of the day, in case it sees something private. Perhaps the bedroom would be off-limits at night. A second type of technical solution relies on detection or recognition: certain objects, classes of objects, or people might be designated as needing to be blurred out whenever they are detected. For example, maybe a concern about the robot seeing secret documents could be addressed by blurring out anything detected as text. For security the blurring should happen at the lowest level possible, such as in the firmware of the camera where it cannot be accessed by the robot's operating system. Finally, a third class of technical solutions would take a "privacy by design" approach by mounting the robot's camera too low to see the tops of tables and desks where documents tend to be. There is a tradeoff here, though, since it might need to see things up high to do certain tasks. For

future work we recommend to implement each of these solutions on a real robot and do exploratory testing to identify broad problems.

7.2.1.3 Data Preferences

In many situations, users will be asked to tell the robot their privacy preferences: which data to collect, where to go, and what to filter. Users might want to adjust these preferences on the fly as issues arise or situations change, perhaps even modifying or deleting data that has already been collected. In future work, interfaces for specifying visual or spatial privacy preferences should be designed. Such interfaces could be like those evaluated by Rueben et al. [198], or they could use other modalities like spoken dialogue, gestures, or drawing. It will be important for these interfaces to accommodate subtle changes to privacy preferences over time.

A problem arises about how to classify information so the robot knows which information can be shared and with whom. For example, robots might talk to humans or amongst themselves, perhaps to gain more information for learning, and might say things that another user would have liked to keep private. A possible solution could be to use an "opt-in" regime, so the robot won't collect or share certain personal information unless you explicitly give your permission ("opt in" to it). Another approach is for the robot to recognize different contexts automatically and respect the different privacy rules for each context as well as for moving between contexts—i.e., to use Nissenbaum's idea of contextual integrity [172]. The robot would just need to be able to detect each context without collecting any

information that is private in that context. Building this sort of Nissenbaumian framework would be completely new research for human-robot interaction.

7.2.1.4 Personalization, Learning, and Inference

Social robots might observe our behavior and adapt to it over time. This behavioral data might even be collected in a private (e.g., home) setting. Perhaps a family's daily routine would be observed by the robot and examined for patterns using machine learning techniques. Since the interpretations of the private information collected by the robot are not only stored but embodied—e.g., if the robot begins to mimic that daily routine—a violation of the users' privacy could happen. Traits and characteristics of family members or even of the whole family might be visible in the behavior of the robot in front of visitors or strangers. Maybe the robot does something impolite during a dinner party that it had clearly learned from a family member. Future work could enable robots to understand which behaviors that were learned in one context are inappropriate to display in other contexts because of the inferences people could make. Alternatively, the robot could simply revert to its default, unpersonalized behaviors around strangers.

7.2.2 Theme 2 of 7: Manipulation and Deception

Humans use different behaviors and personae depending on whether they are with family, friends, coworkers, or strangers. Like humans a robot should be able to

adapt its personality to different people and situations to improve its relationships. For example, a social robot may act more familiarly with a friend than with a stranger. But this also opens the door for robots to be manipulative or deceptive by pretending to be something they are not. This is a privacy risk inasmuch as it could give the robots supervisors access to users' personal information. The question is: how is this process different (if at all) for manipulative robots than for manipulative humans?

7.2.2.1 Social Engineering using Robots

It is possible that robots could be used to trick, con, or dupe humans using social engineering techniques [244, 14]. For example, Booth et al. [34] showed that robots could use social engineering techniques to sneak into a campus building. Other kinds of attacks we foresee include hijacking robots and using them to surveil an area that can be attacked or robbed in the future, or just disabling the robot's cameras so a crime can occur without being recorded. Research on social robots is still very young, so lots of work is still needed to understand which social engineering techniques could be performed by robots, perhaps including new ones that humans cannot perform.

7.2.2.2 Security Vulnerabilities

Fast-paced development of complex systems can leave behind security holes. If a malicious person hacks into a robot there are particularly severe privacy risks because of both the sensors onboard and the robot's ability to move, perhaps in a private setting. There have been some recent examples of security vulnerabilities in robots being exploited to violate privacy. A Hello Barbie doll was hijacked [86] and turned into a surveillance device to spy on children—it recorded the conversations and sends them via WiFi to the internet. Similarly, teddy bears could be used to spy in houses or talk to children [77]. Research in this area might include designing architectures that are especially careful to protect privacy-relevant features of the robot such as sensor feeds, stored data, and motor control. Researchers should also look into limiting the data that gets stored or transmitted, or transmitting only the information that is necessary instead of full-fidelity, raw data [47].

7.2.2.3 Education

One way to mitigate the risk of maleficent deception would be to promote education and experience with robots over time. Robots are still relatively new to our society; when computers were still new, old types of scams were quite successful, like the Nigerian Prince advance fee scam. Now, through education and media, the success rate of those types of scams is relatively low—perhaps education and awareness-raising can do the same for scams that use robots. Even the simple lesson that

robots are built around computers and therefore could have security vulnerabilities could prevent a lot of harm.

People should also know their rights. In the EU, all robots that collect personal data will fall under the General Data Protection Regulation (GDPR)[8], which mandates that users be informed of how their data will be used before they consent to data collection. In the US, a deceptive robot could be regulated by the Federal Trade Commission (FTC) with similar requirements [100]. Educational programs about privacy law and risk management have been developed—e.g., by Daniel J. Solove of TeachPrivacy[9]—but making these trainings specific to robots is an almost untouched research area.

7.2.3 Theme 3 of 7: Trust

The topic of trust is closely related to privacy concerns. Richards and Hartzog [190] have argued that we should not just focus on privacy harms, but also on how privacy can enable trust. Rousseau et al. [194] have defined trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviors of another" (p. 395). Trust has several sub-dimensions, including one's general disposition towards trusting people (e.g., strangers), trust of institutions, trust towards individual people based on what one knows about them, and the willingness to engage in a transaction based on that

[8] https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

[9] https://teachprivacy.com

trust [161]. With regards to robots, research has started to investigate user trust through qualitative, quantitative and experimental methods. Hancock et al. [97] review much of this research in their meta-analysis of which factors affect trust in human-robot interaction.

Future research should continue looking for links between privacy concerns and trust in human-robot interaction because trust is so important for a successful interaction. In addition, researchers should study the factors that affect users' trust that a privacy-protecting system is actually protecting their privacy. Whether it's trust of the software, the manufacturer, or the robot itself, the privacy protection system is not nearly as useful if people aren't confident that their privacy is actually being protected. Finally, future work should study how privacy concerns intersect with trusting relationships like between a patient and a (robot) doctor or if the robot is treated as a family member and is expected not to gossip about what happens at home.

7.2.4 Theme 4 of 7: Blame and Transparency

One of the core questions in privacy-sensitive robotics is whom to blame when privacy violations occur. This is especially difficult when the robot's behavior is at least partially influenced by a remote operator or supervisor, or the robot is capable of learning, or has behavior that is difficult to predict for some other reason. Here we consider several parties that could be blamed: designers/manufacturers, owners, distributors/controllers, and the robot itself.

1. *Designers or manufacturers* may be blamed for having provided improper or faulty privacy protections on their (learning) robot. Alternatively, since the manufacturers cannot control the stimuli from which the robot learns, one could argue that a robot designed, produced or programmed in accordance with privacy standards would not make the manufacturer liable for privacy breaches resulting from a robot's learning capacities. Future work should study existing product liability law regimes across different jurisdictions and whether they need to be changed to better handle robots that can learn.

2. *Owners* may be liable for mistakes made by their robots. Robots with learning capabilities can to some extent be compared to children or pets, for which parents or owners are often held liable even when they committed no fault themselves [42]. Could users be taught that they also have a responsibility for "raising" the robot, and could manufacturers use disclaimers to escape liability in this regard? If so, when (if ever) has the robot "grown up", and who is responsible then?

3. Data assembled by a robot may be transferred to a certain *controlling company*, especially whenever owners have consented to this transfer in general terms and conditions. That company may sell your personal data and allow other businesses to use the robot to give you personalized suggestions. As the robot collects that data it is important to clarify to what extent data controlling companies must respect the privacy of certain information and whether robots need to be programmed to only transfer certain kinds of information.

4. Questions remain when and to what extent *the robot itself* is to be considered a moral agent that is liable for its actions. At some point, the robot may have reached the maximum of its learning abilities and the question can then be asked whether owners are still liable for privacy-breaching actions at that point.

It is also important to be *transparent* about who is actually watching you and recording or using your data, especially since the answer will probably affect your level of privacy concern. Are people more (or less) conscious of their privacy if they know who is behind the robot interpreting the data? When a robot is autonomous, should information about its programming be given? More ethical and sociological research is needed here, setting up studies about how people make privacy-related attributions in instances of shared control over robots. The results of those studies can help us develop features that increase transparency about robotic systems that handle personal information.

7.2.5 Theme 5 of 7: Legal

7.2.5.1 Robots as Persons or Family Members

As social robots enter the home they could be treated like members of the family in a similar way to how pets are sometimes treated. Over time, the humans might build up trust towards the robot so that they speak and act freely around it, expecting that it will respect their privacy at home. This trust is a privacy

risk if robots are subject to search by law enforcement with a proper warrant, access by law enforcement without a warrant under the third party doctrine[10], or perhaps robots will even be mandatory reporters of any signs of domestic abuse. This is especially problematic because robots often have the sensing capabilities (e.g., cameras and microphones, plus the ability to move around the house) to assemble a very detailed picture of home life. One solution would be to consider the robot a close family member with evidentiary privilege in court—i.e., the robot would not have to testify[11] against the family. This might make sense as a way to avoid chilling effects on expression and behavior inside the home—Australia grants evidentiary privilege to the parent-child relationship for similar reasons [68]. A more extreme solution would be to consider robots as "electronic persons" as proposed by the European Parliament[12] instead of as objects of search, forcing police to go through some sort of interrogation process instead of having access to *everything* stored on the robot. Research by legal scholars is needed to work out the details for each of these options and to consider which would be the most reasonable.

[10]In the US, one can have no reasonable expectation of privacy for information that has been given to third parties such as banks or internet service providers ... or, perhaps, robots.

[11]Here, "testifying" might mean having its hard drive accessed and the contents summarized by an analyst.

[12]See European Parliament, resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))

7.2.5.2 Regulating Robotics Companies

Should legal obligations be imposed on businesses to include privacy-sensitive features in their robotic products? Is it legal to compel manufacturers to build in privacy features limiting recording and storage capacities of robots to protect the privacy of their owners? Privacy protection does not always align with business interests because the features that customers want for their robots often require collecting, processing and storing personal information. In the European Union, the right to data protection and especially the right to be forgotten may impact the way robots are designed—more research is needed to determine how.

An alternative to government regulation would be to incentivize businesses to self-regulate. One possible solution would be through a technical standard created by the International Standards Organization (ISO). Standards are adopted without government oversight and could bring confidence that a product is safe, reliable and of sufficient quality. Research in this area could determine the feasibility and usefulness of creating a standard for privacy-sensitive robots.

7.2.5.3 Privacy Education for Users

Should there be a legal obligation to provide privacy education to robot users? If so, who should be responsible for providing it? Public education plans could mandate teachings on what to say and what not to say to robots, and could be applied to children from a young age. Solove and Hartzog have argued that the US Federal Trade Commission (FTC) and its equivalents in other countries may already have

the powers to intervene in this field [224]. Research questions arise about what ways consumer protection agencies can play a role, whether they have sufficient intervention powers or whether their actions should be accompanied by a more developed legal framework. The European Union has proposed the establishment of a robotics agency playing this role[13].

7.2.6 Theme 6 of 7: Private Domains

Some domains have special privacy concerns that warrant their own, additional research efforts. Here we survey three such domains and the special challenges of deploying robots in each.

7.2.6.1 Robotic Surgeons

Use of robotic surgeons such as the daVinci surgical system allows surgeons to perform very precise operations inside the body [219]. Although the robotic surgeons being used right now are teleoperated, advancements in computer vision and robotics may soon make it possible for the robots to carry out some repetitive surgical tasks autonomously. One way to carry out such delicate tasks would involve collecting large sets of sensitive, personal data, including images of a person's naked body. It is essential to have very clear imagery of the environment during surgery, so it becomes difficult to use privacy filters like those proposed by Hubers et al. [108] and Butler et al. [40]. Utility will supercede privacy during surgery

[13]European Parliament, Resolution of 16 February 2017, para 15 and onwards.

since human lives are at risk. It is important, though, to use very secure network connections for communication between the surgeon and the robot, as people could break in and steal this highly sensitive data. It would be even worse if someone were able to get control of the robot and send commands that cause it to harm the patient.

7.2.6.2 Robotic Nurses and Caretakers

Robotic nurses and caretakers could provide a cheaper alternative to human caretakers for providing care in hospitals or even in homes. Robots could help people with disabilities, or older adults who want to age in place [219]. To operate, these robots will need to collect private information such as images of the person and of the environment, maps of the house, and medical information. This raises plenty of concerns even outside of healthcare applications, but robotic nurses and caretakers might be treating people who are especially vulnerable to privacy harms: people who might not be strong enough to resist invasive behaviors or mentally sound enough to understand what is happening. This could include infants and children. Plus, patients might be embarrassed about their appearance or the medical procedures they undergo, so it is more likely that these robots will capture information that should be kept private. Note that many of these same concerns apply to robotic toys designed for children.

7.2.6.3 Robots in the Home

Denning et al. [59] conducted a study exposing security vulnerabilities on three household robots: the WowWee Rovio, the Erector Spykee, and the WowWee RoboSapien V2. Domestic robots like the Roomba and Jibo offer great utility for homes, but with their wide range of sensors they also get access to a lot of private information. For example, the Roomba 960 vacuum robot can build a complete map of a home and can interface with Alexa and the Google Assistant[14]. This integrated network of home devices could be useful, but also more vulnerable to privacy and security threats. What if the companies that manufacture these robots were to sell this private data to other companies that you don't trust? There is also ambiguity about the use of this data; for example, what happens when the previous owner of a house decides to sell their Roomba's map of the house to someone else without the consent of the current owner?

7.2.7 Theme 7 of 7: Theories and Perceptions of Privacy

We need to understand privacy if we are going to build privacy-sensitive robots. Several systematic reviews have been published recently about privacy research in the social sciences [21, 23, 129] and more particularly in communication and information systems [222]. This section is dedicated to the areas of general privacy research (i.e., no robots required) that will be most important and inspirational to privacy-sensitive robotics.

[14]http://www.irobot.com/For-the-Home/Vacuuming/Roomba.aspx

7.2.7.1 Theories or Models of Human Privacy

In addition to economic and legal scholarship, privacy as a social norm has been an important topic in sociology and communication. Prominent theories developed in these disciplines include communication privacy management theory [181], privacy as contextual integrity [172], networked privacy [158], and privacy by design [47], although the latter is also connected to more technical disciplines.

These theories can help us build frameworks for privacy protection. For example, a framework inspired by Nissenbaum's [172] contextual integrity would use the idea of appropriateness and distribution rules in different contexts, whereas communication privacy management theory [181] might inspire the idea of privacy boundaries being modeled as "thick" vs. "thin" depending on how bad it would be if that boundary were breached. Robots equipped with theory-inspired privacy protection frameworks could then be used to test those theories via user studies. This sort of work should promote collaborations by HRI researchers with privacy researchers in law, philosophy, the social sciences, privacy and security experts, and computer scientists.

7.2.7.2 The Subjective Value of Privacy

How much do individuals value privacy relative to other things such as convenience or safety? Research on the *privacy paradox* has shown that in many scenarios users value privacy but are quick to give it up for short-term monetary or social rewards. Dinev and Hart [61] postulate that users perform a privacy calculus by consciously

weighing the benefits of a transaction or service against its privacy risks. Privacy is in that sense like a commodity [46] that can be traded in against a benefit such as access to an affordable ride through Uber or to potential dating partners through Tinder. Despite this, people—including younger people—still value their privacy: survey results have shown little difference between adults and minors in their concern about privacy [158].

How much people value privacy can influence how willing they are to engage with robots that collect personal information. Very little research has studied the subjective value of privacy with regards to robots and their usefulness—a rare example is the study by Butler et al. [40] that first used the phrase "privacy-utility tradeoff" in HRI.

We would recommend future studies in HRI to look at the value of privacy from different methodological standpoints. Ethnographic and observational studies in natural settings could look at how users of social robots trade off privacy for other things like dependability or personalization. Making these studies longitudinal could complement the many studies done on the privacy paradox, which barely look at developmental trajectories over a longer period of time. Also, theoretical approaches such as actor-network theory (ANT) and science and technology studies (STS) could be used to look at how robot engineers and manufacturers build their systems so as to encourage users to prioritize certain values. Similarly to the field experiment by Beresford et al. [31], different "invasiveness" scenarios (a robot collecting more data or accessing more personal rooms) could be combined with

different utility levels (a robot offering more or less useful services to an individual) to test the privacy-utility tradeoff.

7.3 Suggested Collaborations

It should be clear from the diversity of the seven themes presented above that privacy-sensitive robotics research will require a lot of collaboration. We see four different types of collaboration that will be needed: expertise from other disciplines, implementation in different application areas, synergies with related HRI research areas, and working with industry partners to understand all the practical challenges of real-world deployment. This section will give examples of all four types of collaborations.

7.3.1 Collaborations with Experts from other Disciplines

- We need **privacy and security** experts to build architectures that prioritize protecting personal information—and to tell us where vulnerabilities still exist.

- **User experience (UX) and user interface (UI) designers** need to create usable ways to figure out users privacy preferences to be enforced on robots. Experts in **human factors and ergonomics** could contribute more broadly to evaluations of designs for protecting privacy.

- To understand human behavior we will need **social scientists** such as psy-

chologists, sociologists, anthropologists, and economists. They can draw from theories that have already been formulated and tested, and help test whether they still apply in new types of interactions (e.g., with robots).

- We need **lawyers and policymakers** who understand the laws and regulations about privacy and technology. This is especially true when working across borders, such as between the US and EU legal systems.
- The field of **information science** would be helpful for modeling aspects of information privacy in particular, such as the way big data could be used to make inferences about sensitive, personal data from a few public observations.
- Lastly, we will need to understand privacy better—**privacy scholars** from fields like psychology, sociology, communication science, and law need to study how people think about privacy and to form coherent theories and taxonomies.

7.3.2 Collaborations with Experts who Work in Specific Application Areas

- Privacy-sensitive robotics research should be of interest to people who develop **robots for private domains** such as homes and hospitals. Within these domains there are even more private subdomains such as restrooms and bedrooms.
- Other domains could be implicated because of their secrecy, such as **robots**

for industry or the military. Companies might have proprietary information to protect, and military operations are often classified.

- **Social robots** will also need privacy protections. Robots that can physically touch people need to be careful of where and how they touch. Also, some robots encourage people to form strong social bonds with them, which is also risky if the robot can collect data and send it elsewhere.

- Privacy-sensitive robotics researchers should also reach out to people who develop **robots for public spaces** to address potential concerns about surveillance and chilling effects on moving and speaking freely.

7.3.3 Collaborations with Experts in Related HRI Research Areas

There is synergy to be exploited between privacy-sensitive robotics and some other areas of HRI research. For example, privacy research could benefit from findings about concepts like **trust, acceptance, transparency, explainability, theory of mind, and common ground**. The broad areas of **design, machine learning, object recognition, and navigation** are also involved, as are some more specific efforts towards **personalization and automated ethical reasoning**.

7.3.4 Collaborations with Experts from Industry

Privacy-sensitive robotics research will also benefit in multiple ways from collaborations with people from industry. First of all, companies conduct **market re-**

search to understand potential customers: who they are and what they want (and don't want). This is followed by developing **experience with how users interact with the robot** as robots are tested with users, sold to users, and then maintained via technical support. Companies usually place an **emphasis on user acceptance of and satisfaction with the robot** throughout this process. Once robots are being sold and used, companies are able to observe **larger-scale and longer deployments**, often in diverse settings. Through providing technical support services for these robots, companies can learn **the types of problems that can occur in the field**.

8 Conclusion

8.1 Summary of Contributions

This dissertation has presented a program of research designed to help launch a new research area to address the need for privacy research in human-robot interaction. We call it "privacy-sensitive robotics." We have already published an extensive review of the relevant privacy scholarship as well as a list of tools from computer science and robotics that could be used for privacy protection [195]. We have also published reports on two of our experimental contributions [198, 199]. The first tested whether physical markers were better than a GUI for specifying privacy preferences, and found the answer to be a trade-off. Specifically, the GUI interface was faster and easier to use in the study context, whereas the physical marker interface seemed to aid users' memory, thereby lending itself to higher-stakes situations. The second published study was a series of four online surveys that consistently found a large effect of framing on privacy concerns about a telepresence robot in a home setting. This highlighted to designers the importance of learning about how to set the frame for an interaction and how it influences privacy judgments.

The report for our third study is being prepared for submission to a journal. The study was a six week, in-the-wild deployment of a novel "mobile shoe rack"

robot to study mental model formation. Privacy-relevant findings included: (1) participants formed beliefs about the robot's sensing capabilities even though we designed them to be ambiguous; and (2) all six of our participants were fooled into thinking the robot was autonomous.

Finally, we have made recommendations for the future of privacy-sensitive robotics. We laid out a strategic roadmap, including key steps like developing measurement tools for privacy concerns as well as community activities like announcing challenge problems to stimulate research. We also presented seven theme areas in privacy-sensitive robotics—data privacy, manipulation and deception, trust, blame and transparency, legal concerns, private domains, and privacy scholarship—along with new research ideas for each. Lastly, we recommended a lot of different collaborations: with other human-robot interaction researchers, people from other academic disciplines, and people from industry. Working with people from all these areas is part of our due diligence for addressing privacy issues in human-robot interaction.

8.2 A Call to Action

We have seen that privacy is a key value to uphold in any human society. It's a big, multifaceted concept that goes beyond personal information to touch our freedom, relationships, and personal growth. Robots promise to make our lives better, but their actions will become increasingly privacy-relevant as they become

more ubiquitous and social. A call to action is in order considering the current sluggishness of progress in privacy-sensitive robotics research.

We believe there is an opportunity in the next few years to make a positive difference if we take strategic action. We see two possibilities for the near future. First, to a certain extent people will bend but not break—i.e., people's privacy expectations will lower and it will become harder to keep a higher level of privacy as society changes. Perhaps some of this will be ethically neutral, but other parts will not—now is the time to fight back against giving in where we should remain firm. Second, there could also be privacy disasters in the future—events like security breaches or exposés of long abuses, events that really harm people and generate a public reaction in favor of more privacy protection. This backlash could be healthy for a culture that does not value privacy enough, but we might be able to avoid the disasters altogether if we improve our culture now and make wise decisions about robotics technologies.

We therefore urge our readers to consider contributing to privacy-sensitive robotics work. This is not just a call for HRI researchers to study privacy—we also encourage privacy and security researchers to consider studying robots. Privacy-sensitive robotics researchers should push for the development of this new research area. This will involve the usual means: e.g., establishing a community by organizing workshops, reaching out to other disciplines and to industry for collaborations, and charging forward on strategic research goals—for example, the ones we have recommended. HRI researchers who study privacy-relevant domains should link up with us and begin to consider privacy in their research. Prudent ac-

tion now from a critical mass of different actors could greatly improve the outcome of introducing robots into society.

Bibliography

[1] U.S. Constitution. Amend. IV., 1791.

[2] U.S. Constitution. Amend. I., 1791.

[3] Design for accessibility: A cultural administrator's handbook. 2003.

[4] Online Etymology Dictionary, 2015. URL `http://www.etymonline.com/`.

[5] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, (1):26–33, 2005.

[6] Alessandro Acquisti, Leslie K. John, and George Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.

[7] Anita Allen. Privacy in Health Care. In Warren Thomas Reich, editor, *Encylopedia of Bioethics*, volume 4. Simon & Schuster, New York, 1995.

[8] Anita Allen. Coercing privacy. *Wm. & Mary L. Rev.*, 40:723, 1998.

[9] Anita Allen. Privacy and Medicine. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2011 edition, 2011.

[10] Irwin Altman. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* Brooks/Cole Publishing Company, Monterey, CA, 1975.

[11] Irwin Altman. Privacy Regulation: Culturally Universal or Culturally Specific? *Journal of Social Issues*, 33(3):66–84, 1977.

[12] Mary Applegate and Janice M. Morse. Personal Privacy and Interaction Patterns in a Nursing Home. *Journal of Aging Studies*, 8(4):413–434, 1994.

[13] Aristotle. *Politics*. NuVision Publications, LLC, 2004.

[14] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to its Recommendations or Gamble? *IEEE Robotics and Automation Letters*, ([in press]), 2018.

[15] Hajime Asama, Koichi Ozaki, Hiroaki Itakura, Akihiro Matsumoto, Yoshiki Ishida, and Isao Endo. Collision avoidance among multiple mobile robots based on rules and communication. In *Proceedings of IEEE/RSJ International Workshop on Intelligent Robots and Systems*, pages 1215–1220. IEEE, 1991.

[16] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805, October 2010. ISSN 1389-1286. doi: 10.1016/j.comnet.2010.05.010.

[17] Lisa Austin. Privacy and the Question of Technology. *Law and Philosophy*, 22(2):119–166, 2003.

[18] Atta Badii, Mathieu Einig, Tomas Piatrik, and others. Overview of the MediaEval 2013 Visual Privacy Task. In *MediaEval*, 2013.

[19] Kevin S. Bankston and Amie Stepanovich. When robot eyes are watching you: the law & policy of automated communications surveillance. In *Proceedings of We Robot 2014*, University of Miami, 2014. Draft.

[20] Chelsea Barabas, Christopher Bavitz, J. Nathan Matias, Cecillia Xie, and Jack Xu. Legal and ethical issues in the use of telepresence robots: best practices and toolkit. In *Proceedings of We Robot 2015*, University of Washington, 2015. Draft.

[21] Susanne Barth and Menno D T De Jong. The Privacy Paradox—Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior—A Systematic Literature Review. *Telematics and Informatics*, 34:1038–1058, 2017. ISSN 07365853. doi: 10.1016/j.tele.2017.04.013.

[22] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1:71–81, January 2009. ISSN 1875-4791, 1875-4805.

[23] Lemi Baruh, Ekin Secinti, and Zeynep Cemalcilar. Online Privacy Concerns and Privacy Management: A Meta-Analytical Review. *Journal of Communication*, 67(1):26–53, 2017. ISSN 14602466. doi: 10.1111/jcom.12276.

[24] Gregory Bateson. A theory of play and fantasy. *Psychiatric research reports*, 1955.

[25] Franklin D. Becker and Clara Mayo. Delineating personal distance and territoriality. *Environment and Behavior*, 1971.

[26] Jenay M. Beer and Leila Takayama. Mobile Remote Presence Systems for Older Adults: Acceptance, Benefits, and Concerns. In *Proceedings of the 6th International Conference on Human-robot Interaction*, HRI '11, pages 19–26, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0561-7. doi: 10.1145/1957656.1957665.

[27] Victoria Bellotti and Abigail Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 1317 September 1993, Milan, Italy ECSCW93*, pages 77–92. Springer, 1993.

[28] Gary Bente, Sabine Rüggenberg, and N. C. Krämer. Social presence and interpersonal trust in avatar-based, collaborative net-communications. In *7th Annual International Workshop on Presence*, 2004.

[29] Gary Bente, Felix Eschenburg, and Lisa Aelker. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain*, pages 207–14, 2007.

[30] Bettina Berendt, Oliver Gnther, and Sarah Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 48(4):101–106, 2005.

[31] Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1):25–27, 2012. ISSN 01651765. doi: 10.1016/j.econlet.2012.04.077.

[32] Edward J Bloustein. Privacy as an aspect of human dignity: An answer to Dean Prosser. In Ferdinand David Schoeman, editor, *Philosophical dimensions of privacy: An anthology*. Cambridge University Press, 1964.

[33] Patrick Boissy, Hlne Corriveau, Franois Michaud, Daniel Labont, and Marie-Pier Royer. A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects. *Journal of Telemedicine and Telecare*, 13(2):79–84, 2007.

[34] Serena Booth, James Tompkin, Hanspeter Pfister, James Waldo, Krzysztof Gajos, and Radhika Nagpal. Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security. In *Proceedings of the twelfth annual ACM/IEEE international conference on Human-Robot Interaction - HRI '17*, pages 426–434, 2017. ISBN 978-1-4503-4336-7. doi: dx.doi.org/10.1145/2909824.3020211.

[35] Michael Boyle, Christopher Edwards, and Saul Greenberg. The effects of filtered video on awareness and privacy. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 1–10. ACM, 2000.

[36] Michael Boyle, Carman Neustaedter, and Saul Greenberg. Privacy factors in video-based media spaces. In *Media Space 20+ Years of Mediated Life*, pages 97–122. Springer, 2009.

[37] Tom Buchanan, Carina Paine, Adam N. Joinson, and Ulf-Dietrich Reips. Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology*, 58(2):157–165, 2007. ISSN 1532-2882.

[38] Aurlie Bugeau, Marcelo Bertalmo, Vicent Caselles, and Guillermo Sapiro. A comprehensive framework for image inpainting. *Image Processing, IEEE Transactions on*, 19(10):2634–2645, 2010.

[39] Judee Burgoon. Privacy and communication. In Michael Burgoon, editor, *Communication Yearbook 6*, number 6. Routledge, 1982.

[40] Dan Butler, Justin Huang, Franziska Roesner, and Maya Cakmak. The Privacy-Utility Tradeoff for Remotely Teleoperated Robots. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Portland, OR, 2015.

[41] John Travis Butler and Arvin Agah. Psychological effects of behavior patterns of a mobile personal robot. *Autonomous Robots*, 10(2):185–202, 2001.

[42] Erdem Buyuksagis and H. Van Boom Willem. Strict liability in European codification. Torn between objects, activities and their risks. *Geo. J. Int'l L. 44*, page 609, 2012.

[43] Kelly Caine, Selma Sabanovic, and Mary Carter. The Effect of Monitoring by Cameras and Robots on the Privacy Enhancing Behaviors of Older Adults. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 343–350, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1063-5. doi: 10.1145/2157689.2157807.

[44] M Ryan Calo. The drone as a privacy catalyst. *Stan. L. Rev. Online*, 64:29, 2011.

[45] Ryan Calo. Robots and privacy. *Robot Ethics: The Ethical and Social Implications of Robotics, Patrick Lin, George Bekey, and Keith Abney, eds., Cambridge: MIT Press*, 2010.

[46] John Edward Campbell and Matt Carlson. Panopticon.com: Online Surveillance and the Commodification of Privacy. *Journal of Broadcasting & Electronic Media*, 46(4):586–606, 2002. ISSN 0883-8151.

[47] Ann Cavoukian. *Privacy by Design: The 7 Foundational Principles.* Information and Privacy Commissioner of Ontario, 2016. http://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/.

[48] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. Perceived robot capability. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 541–548. IEEE, 2015.

[49] Tiffany L. Chen, Matei Ciocarlie, Steve Cousins, Phillip M. Grice, Kelsey Hawkins, Kaijen Hsiao, Charles C. Kemp, Chih-Hung King, Daniel A. Lazewatsky, Adam E. Leeper, Hai Nguyen, Andreas Paepcke, Caroline Pantofaru, William D. Smart, and Leila Takayama. Robots for Humanity: Using Assistive Robotics to Empower People with Disabilities. *IEEE Robotics and Automation Magazine*, 20(1):30–39, March 2013.

[50] S.-C.S. Cheung, Mahalingam Vijay Venkatesh, J.K. Paruchuri, Jian Zhao, and Thinh Nguyen. Protecting and managing privacy information in video surveillance systems. In *Protecting Privacy in Video Surveillance*, pages 11–33. Springer, 2009.

[51] S. S. Cheung, Jian Zhao, and M. Vijay Venkatesh. Efficient object-based video inpainting. In *Image Processing, 2006 IEEE International Conference on*, pages 705–708. IEEE, 2006.

[52] Forrester Cole, Douglas DeCarlo, Adam Finkelstein, Kenrick Kin, R. Keith Morley, and Anthony Santella. Directing Gaze in 3d Models with Stylized Focus. *Rendering Techniques*, 2006:17th, 2006.

[53] James L Crowley, Joëlle Coutaz, and François Bérard. Things That See. *Communications of the ACM*, 43(3):54–ff, 2000.

[54] Kate Darling. "Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In *Proceedings of We Robot 2015*, University of Washington, 2015. Available at SSRN: http://ssrn.com/abstract=2588669 or http://dx.doi.org/10.2139/ssrn.2588669.

[55] Maartje De Graaf, Somaya Ben Allouch, and Jan Van Dijk. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 224–233. ACM, 2017.

[56] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 769–776. ACM, 2002.

[57] Doug DeCarlo, Adam Finkelstein, Szymon Rusinkiewicz, and Anthony Santella. Suggestive contours for conveying shape. *ACM Transactions on Graphics (TOG)*, 22(3):848–855, 2003.

[58] Judith DeCew. Privacy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition, 2013.

[59] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R. Smith, and Tadayoshi Kohno. A spotlight on security and privacy risks with future

household robots: attacks and lessons. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 105–114. ACM, 2009.

[60] Robert F DeVellis. *Scale development: Theory and applications*. Sage publications, fourth edition, 2016.

[61] Tamara Dinev and Paul Hart. An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1):61–80, 2006. ISSN 15265536. doi: 10.1287/isre.1060.0080.

[62] A.D. Dragan and S.S. Srinivasa. Generating legible motion. In *Proceedings of Robotics: Science and Systems (R:SS)*, 2013.

[63] Julian J. Edney and Michael A. Buda. Distinguishing territoriality and privacy: Two studies. *Human Ecology*, 4(4):283–296, 1976.

[64] Felix Endres, Jrgen Hess, Nikolas Engelhard, Jrgen Sturm, Daniel Cremers, and Wolfram Burgard. An evaluation of the RGB-D SLAM system. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1691–1696. IEEE, 2012.

[65] Ádám Erdélyi, Tibor Barat, Patrick Valet, Thomas Winkler, and Bernhard Rinner. Adaptive cartooning for privacy protection in camera networks. In *11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 44–49. IEEE, 2014.

[66] Ádám Erdélyi, Thomas Winkler, and Bernhard Rinner. Multi-level cartooning for context-aware privacy protection in visual sensor networks. In *MediaEval*, 2014.

[67] Clifton A Ericson et al. *Hazard analysis techniques for system safety*. John Wiley & Sons, 2015.

[68] Hillary Farber. To testify or not to testify: A comparative analysis of Australian and American approaches to a parent-child testimonial exemption. *Tex. Int'l LJ*, 46:109, 2010.

[69] Ylva Fernaeus, Maria Håkansson, Mattias Jacobsson, and Sara Ljungblad. How do you play with a robotic toy animal?: a long-term study of pleo. In *Proceedings of the 9th international Conference on interaction Design and Children*, pages 39–48. ACM, 2010.

[70] Julia Fink, Valérie Bauwens, Frédéric Kaplan, and Pierre Dillenbourg. Living with a vacuum cleaning robot. *International Journal of Social Robotics*, 5 (3):389–408, 2013.

[71] Jan Fischer, Douglas Cunningham, Dirk Bartz, Christian Wallraven, Heinrich Beulthoff, and Wolfgang Strasser. Measuring the discernability of virtual objects in conventional and stylized augmented reality. In *12th Eurographics Symposium on Virtual Environments, Lisbon, Portugal, May 8th-10th, 2006*, page 53. Transaction Publishers, 2006.

[72] K. Fischer, B. Soto, C. Pantofaru, and L. Takayama. Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance. In *2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 999–1005, August 2014.

[73] Kerstin Fischer. Interpersonal Variation in Understanding Robots As Social Actors. In *Proceedings of the 6th International Conference on Human-robot Interaction*, HRI '11, pages 53–60, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0561-7.

[74] Robert S Fish, Robert E Kraut, Robert W Root, and Ronald E Rice. Evaluating video as a technology for informal communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 37–48. ACM, 1992.

[75] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003.

[76] David A Forsyth, Margaret Fleck, and Chris Bregler. Finding naked people. *International Journal of Computer Vision*, 1996.

[77] Lorenzo Franceschi-Bicchierai. How This Internet of Things Stuffed Animal Can Be Remotely Turned Into a Spy Device, 2017. https://motherboard.vice.com/en_us/article/qkm48b/how-this-internet-of-things-teddy-bear-can-be-remotely-turned-into-a-spy-device.

[78] Charles Fried. An anatomy of values. *Cambridge, Mass*, 1970.

[79] Batya Friedman, Peter H. Kahn Jr, Jennifer Hagman, Rachel L. Severson, and Brian Gill. The Watcher and the Watched: Social Judgments about Privacy in a Public Place. In *Media Space 20+ Years of Mediated Life*, pages 145–176. Springer, 2009.

[80] Batya Friedman, Peter H Kahn, Alan Borning, and Alina Huldtgren. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*, pages 55–95. Springer, 2013.

[81] Michael R Furr and Verne R Bacharach. *Psychometrics. An introduction*. Sage Publications, Thousand Oaks, CA, 2008.

[82] Susan R Fussell, Sara Kiesler, Leslie D Setlock, and Victoria Yew. How people anthropomorphize robots. In *3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 145–152. IEEE, 2008.

[83] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. (IROS 2005)*, pages 2278–2283, August 2005. doi: 10.1109/IROS.2005.1545511.

[84] Cipriano Galindo, Juan-Antonio Fernndez-Madrigal, Javier Gonzlez, and Alessandro Saffiotti. Robot task planning using semantic maps. *Robotics and Autonomous Systems*, 56(11):955–966, 2008.

[85] Timothy Gerstner, Doug DeCarlo, Marc Alexa, Adam Finkelstein, Yotam Gingold, and Andrew Nealen. Pixelated image abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, pages 29–36. Eurographics Association, 2012.

[86] S Gibbs. Hackers can hijack Wi-Fi Hello Barbie to spy on your children, 2015. https://www.theguardian.com/technology/2015/nov/26/hackers-can-hijack-wi-fi-hello-barbie-to-spy-on-your-children.

[87] Michael J Gill, William B Swann Jr, and David H Silvera. On the genesis of confidence. *Journal of Personality and Social Psychology*, 75(5):1101, 1998.

[88] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin

Snipes, Alan C Schultz, et al. Designing robots for long-term social interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1338–1343. IEEE, 2005.

[89] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. Natural Person-following Behavior for Social Robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, HRI '07, pages 17–24, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-617-2. doi: 10.1145/1228716.1228720.

[90] Amy Ashurst Gooch and Peter Willemsen. Evaluating space perception in NPR immersive environments. In *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, pages 105–110. ACM, 2002.

[91] Michael A. Goodrich and Alan C. Schultz. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3):203–275, 2007.

[92] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. Affective personalization of a social robot tutor for children's second language skills. In *AAAI*, pages 3951–3957, 2016.

[93] Lucian Cosmin Goron, Zoltan-Csaba Marton, Dejan Pangercic, Thomas Ruhr, Moritz Tenorth, and Michael Beetz. Autonomous Semantic Mapping for Robots Performing Everyday Manipulation Tasks in Kitchen Environments. In *Proceedings of the International Conference on Robots and Systems (IROS)*, pages 4263–4270, 2011.

[94] Victoria Groom, Vasant Srinivasan, Cindy L. Bethel, Robin Murphy, Lorin Dole, and Clifford Nass. Responses to robot social roles and social role framing. In *International Conference on Collaboration Technologies and Systems (CTS)*, pages 194–203. IEEE, 2011.

[95] Edward T. Hall. *The Hidden Dimension*. Doubleday, Garden City, 1966.

[96] Nick Halper, Mara Mellin, Christoph S. Herrmann, Volker Linneweber, and Thomas Strothotte. Towards an understanding of the psychology of non-photorealistic rendering. In *Computational Visualistics, Media Informatics, and Virtual Communities*, pages 67–78. Springer, 2003.

[97] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527, 2011. ISSN 0018-7208. doi: 10.1177/0018720811417254.

[98] Harris and Associates, Inc. and Alan Westin. E-commerce and privacy: What net users want. *Privacy and American Business, Hackensack, NJ*, 1998.

[99] Woodrow Hartzog. Focus on cyberlaw: Unfair and deceptive robots. *Maryland Law Review*, 74:785–1031, 2015.

[100] Woodrow Neal Hartzog. Et Tu, Android? Regulating Dangerous and Dishonest Robots. *Journal of Human-Robot Interaction*, 5(3):70, 2016. ISSN 2163-0364. doi: 10.5898/JHRI.5.3.Hartzog.

[101] Z. Henkel, C.L. Bethel, R.R. Murphy, and V. Srinivasan. Evaluation of Proxemic Scaling Functions for Social Robotics. *IEEE Transactions on Human-Machine Systems*, 44(3):374–385, June 2014. ISSN 2168-2291. doi: 10.1109/THMS.2014.2304075.

[102] Jan Herling and Wolfgang Broll. Pixmix: A real-time approach to high-quality diminished reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 141–150. IEEE, 2012.

[103] Guy Hoffman, Jodi Forlizzi, Shahar Ayal, Aaron Steinfeld, John Antanitis, Guy Hochman, Eric Hochendoner, and Justin Finkenaur. Robot Presence and Human Honesty: Experimental Evidence. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 181–188, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-2883-8.

[104] Jason I. Hong and James A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 177–189. ACM, 2004.

[105] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. OctoMap: An efficient probabilistic 3d mapping framework based on octrees. *Autonomous Robots*, 34(3):189–206, 2013.

[106] Iris Howley, Takayuki Kanda, Kotaro Hayashi, and Carolyn Rosé. Effects of Social Presence and Social Role on Help-seeking and Learning. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 415–422, New York, NY, USA, 2014. ACM.

[107] Guoqiang Hu, Wee Peng Tay, and Yonggang Wen. Cloud robotics: architecture, challenges and applications. *IEEE network*, 26(3), 2012.

[108] Alexander Hubers, Emily Andrulis, Levi Scott, Tanner Stirrat, Ruonan Zhang, Ross Sowell, Matthew Rueben, Cindy M Grimm, and William D Smart. Using video manipulation to protect privacy in remote presence systems. In *International Conference on Social Robotics*, pages 245–254. Springer, 2015.

[109] Alexander Hubers, Emily Andrulis, Tanner Stirrat, Duc Tran, Ruonan Zhang, Ross Sowell, Cindy M. Grimm, and William D. Smart. Video Manipulation Techniques for the Protection of Privacy in Remote Presence Systems. In *HRI 2015 Extended Abstracts*, Portland, OR, March 2015.

[110] Julie C. Inness. *Privacy, intimacy, and isolation.* Oxford University Press, 1992.

[111] InTouch Health. *InTouch Telemedicine System.* 2012. http://www.intouchhealth.com/.

[112] Kaori Ishii. Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects. *AI & SOCIETY*, pages 1–25, 2017.

[113] Ellen Jacobs. The Need for Privacy and the Application of Privacy to the Day Care Setting. In *Biennial Meeting of the Society for Research in Child Development*, New Orleans, Louisiana, November 1977.

[114] Suman Jana, Arvind Narayanan, and Vitaly Shmatikov. A Scanner Darkly: Protecting user privacy from perceptual applications. In *2013 IEEE Symposium on Security and Privacy (SP)*, pages 349–363. IEEE, 2013.

[115] Cynthia Johnson-George and Walter C. Swap. Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of personality and social psychology*, 43(6):1306, 1982.

[116] Edward E Jones. *Interpersonal perception.* WH Freeman/Times Books/Henry Holt & Co, 1990.

[117] Michiel Joosse, Aziez Sardar, Manja Lohse, and Vanessa Evers. BEHAVE-II: The revised set of measures to assess users attitudinal and behavioral responses to a social robot. *International journal of social robotics*, 5(3): 379–388, 2013.

[118] Peter H. Kahn, Jr., Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Solace Shen, Heather E. Gary, and Jolina H. Ruckert. Will People Keep the Secret of a Humanoid Robot?: Psychological Intimacy in HRI. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 173–180, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-2883-8.

[119] Margot E Kaminski. Robots in the home: What will we have agreed to? *Idaho L. Rev.*, 51:661, 2014.

[120] Margot E Kaminski, Matthew Rueben, William D Smart, and Cindy M Grimm. Averting robot eyes. *Md. L. Rev.*, 76:983, 2016.

[121] Jonathan Karro, Andrew W Dent, and Stephen Farish. Patient perceptions of privacy infringements in an emergency department. *Emergency Medicine Australasia*, 17(2):117–123, April 2005. ISSN 17426731. doi: 10.1111/j.1742-6723.2005.00702.x.

[122] Shin Kato, Sakae Nishiyama, and Jun'ichi Takeno. Coordinating Mobile Robots By Applying Traffic Rules. In *IROS*, volume 92, pages 1535–1541, 1992.

[123] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

[124] Sara Kiesler. Fostering common ground in human-robot interaction. In *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, pages 729–734. IEEE, 2005.

[125] Sara Kiesler and Jennifer Goetz. Mental models of robotic assistants. In *CHI'02 extended abstracts on Human Factors in Computing Systems*, pages 576–577. ACM, 2002.

[126] Hyun Hoi James Kim, Carl Gutwin, and Sriram Subramanian. The magic window: lessons from a year in the life of a co-present media space. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 107–116. ACM, 2007.

[127] Jeffrey Klow, Jordan Proby, Matthew Rueben, Ross T Sowell, Cindy M Grimm, and William D Smart. Privacy, utility, and cognitive load in remote presence systems. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 167–168. ACM, 2017.

[128] Kheng Lee Koay, Dag Sverre Syrdal, Michael L Walters, and Kerstin Dautenhahn. Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *The 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, pages 564–569. IEEE, 2007.

[129] Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122–134, 2017.

[130] Pavel Korshunov and Touradj Ebrahimi. Using face morphing to protect privacy. In *10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 208–213. IEEE, 2013.

[131] Pavel Korshunov, Shuting Cai, and Touradj Ebrahimi. Crowdsourcing approach for evaluation of privacy filters in video surveillance. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*, pages 35–40. ACM, 2012.

[132] Margaret M. Krupp, Matthew Rueben, Cindy M. Grimm, and William D. Smart. A focus group study of privacy concerns about telepresence robots. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO–MAN 2017)*. IEEE, 2017.

[133] Minae Kwon, Malte F Jung, and Ross A Knepper. Human expectations of social robots. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 463–464. IEEE, 2016.

[134] Jan Eric Kyprianidis. Image and Video Abstraction by Multi-scale Anisotropic Kuwahara Filtering. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, pages 55–64, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0907-3. doi: 10.1145/2024676.2024686.

[135] Marianne La France and Clara Mayo. A review of nonverbal behaviors of women and men. *Western Journal of Speech Communication*, 43(2):96–107, 1979.

[136] Marc Langheinrich. Privacy by designprinciples of privacy-aware ubiquitous systems. In *Ubicomp 2001: Ubiquitous Computing*, pages 273–291. Springer, 2001.

[137] Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In *UbiComp 2002: Ubiquitous Computing*, pages 237–245. Springer, 2002.

[138] Steven M. LaValle. *Planning algorithms*. Cambridge University Press, 2006.

[139] Scott Lederer, Anind K. Dey, and Jennifer Mankoff. A conceptual model and a metaphor of everyday privacy in ubiquitous computing environments. Technical Report UCB/CSD-2-1188, Computer Science Division, University of California, Berkeley, 2002.

[140] Scott Lederer, Jennifer Mankoff, and Anind K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 724–725. ACM, 2003.

[141] Min Kyung Lee and Leila Takayama. "Now, I Have a Body": Uses and Social Norms for Mobile Remote Presence in the Workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 33–42, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978950.

[142] Min Kyung Lee, Karen P. Tang, Jodi Forlizzi, and Sara Kiesler. Understanding users' perception of privacy in human-robot interaction. In *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 181–182. IEEE, 2011.

[143] Sau-lai Lee, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2767–2772. IEEE, 2005.

[144] H. Leino-Kilpi, M. Valimaki, T. Dassen, M. Gasull, C. Lemonidou, A. Scott, and M. Arndt. Privacy: A Review of the Literature. *International Journal of Nursing Studies*, 38:663–671, 2001.

[145] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.

[146] Séverin Lemaignan, Julia Fink, Pierre Dillenbourg, and Claire Braboszcz. The cognitive correlates of anthropomorphism. In *HRI '14 Workshop entitled "HRI: a bridge between Robotics and Neuroscience"*, number EPFL-CONF-196441, 2014.

[147] Daniel T. Levin, Stephen S. Killingsworth, and Megan M. Saylor. Concepts About the Capabilities of Computers and Robots: A Test of the Scope of Adults' Theory of Mind. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, HRI '08, pages 57–64, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-017-3.

[148] Daniel T Levin, Caroline Harriott, Natalie A Paul, Tao Zhang, and Julie A Adams. Cognitive dissonance as a measure of reactions to human-robot interaction. *Journal of Human-Robot Interaction*, 2(3):3–17, 2013.

[149] David Levy. *Love and sex with robots: The evolution of human-robot relationships*. Harper, New York, NY, 2009.

[150] Christina Lichtenthäler and Alexandra Kirsch. Towards legible robot navigation—how to increase the intend expressiveness of robot navigation behavior. In *Proceedings of International Conference on Social Robotics – Workshop on Embodied Communication of Goals and Intentions*, 2013.

[151] David Lu and William D. Smart. Towards More Efficient Navigation for Robots and Humans. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[152] David Lu, Daniel B. Allan, and William D. Smart. Tuning Cost Functions for Social Navigation. In *Proceedings of the International Conference on Social Robotics (ICSR)*, 2013.

[153] Jingwan Lu, Pedro V. Sander, and Adam Finkelstein. Interactive painterly stylization of images, videos and 3d animations. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 127–134. ACM, 2010.

[154] Christoph Lutz and Aurelia Tamò. Robocode-ethicists: Privacy-friendly robots, an ethical responsibility of engineers? In *Proceedings of the ACM Web Science Conference*, page 21. ACM, 2015.

[155] Christoph Lutz and Aurelia Tamò. Privacy and healthcare robots—an ANT analysis. In *We Robot 2016: the Fifth Annual Conference on Legal and Policy Issues relating to Robotics*. University of Miami School of Law, 2016. Discussant: Matt Beane, University of California Santa Barbara.

[156] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[157] John Markoff. The Boss Is Robotic, and Rolling Up Behind You. *The New York Times*, September 2010.

[158] Alice E Marwick and Danah Boyd. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7):1051–1067, 2014.

[159] Winter Mason and Siddharth Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1):1–23, 2012.

[160] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona Newell, and Carol O'Sullivan. Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 67–74. ACM, 2008.

[161] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3):334–359, 2002.

[162] Brian Ka-Jun Mok, Stephen Yang, David Sirkin, and Wendy Ju. A place for every tool and every tool in its place: Performing collaborative tasks with interactive robotic drawers. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 700–706. IEEE, 2015.

[163] Adam D. Moore. Intangible Property: Privacy, Power, and Information Control. *American Philosophical Quarterly*, 35(4):365–378, October 1998. ISSN 0003-0481.

[164] Adam D. Moore. Privacy: its meaning and value. *American Philosophical Quarterly*, pages 215–227, 2003.

[165] Jonathan Mumm and Bilge Mutlu. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 331–338. ACM, 2011.

[166] Yuta Nakashima, Tatsuya Koyama, Naokazu Yokoya, and Noboru Babaguchi. Facial expression preserving privacy protection using image melding. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2015.

[167] Yasushi Nakauchi and Reid Simmons. A Social Robot that Stands in Line. *Autonomous Robots*, 12(3):313–324, May 2002. ISSN 0929-5593, 1573-7527. doi: 10.1023/A:1015273816637.

[168] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 72–78, New York, NY, USA, 1994. ACM. ISBN 0-89791-650-6. doi: 10.1145/191666.191703.

[169] P. B. Newell. A cross-cultural comparison of privacy definitions and functions: A systems approach. 1998. ISSN 0272-4944.

[170] Patricia B. Newell. Perspectives on Privacy. *Journal of Environmental Psychology*, 15:87–104, 1995.

[171] Richard E. Nisbett and Timothy D. Wilson. Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3): 231–259, 1977.

[172] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[173] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI & Society*, 20(2):138–150, 2006.

[174] Don Norman. *The design of everyday things: Revised and expanded edition*. Constellation, 2013.

[175] Donald A Norman. Some observations on mental models. In *Mental models*, pages 15–22. Psychology Press, 2014.

[176] Sandra Y. Okita, Victor Ng-Thow-Hing, and Ravi Kiran Sarvadevabhatla. Captain may I?: proxemics study examining factors that influence distance between humanoid robots, children, and adults, during human-robot interaction. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 203–204. ACM, 2012.

[177] Steffi Paepcke and Leila Takayama. Judging a bot by its cover: an experiment on expectation setting for personal robots. In *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 45–52. IEEE, 2010.

[178] Carina Paine, Ulf-Dietrich Reips, Stefan Stieger, Adam Joinson, and Tom Buchanan. Internet users' perceptions of "privacy concerns" and "privacy actions". *International Journal of Human-Computer Studies*, 65:526–536, 2007. ISSN 1071-5819.

[179] Robert Paine. Lappish decisions, partnerships, information management, and sanctions: A nomadic pastoral adaptation. *Ethnology*, 9(1):52–67, 1970.

[180] W. A. Parent. Privacy, Morality, and the Law. *Philosophy & Public Affairs*, 12(4):269–288, October 1983. ISSN 0048-3915.

[181] Sandra Petronio. Communication boundary management: A theoretical model of managing disclosure of private information between marital couples. *Communication Theory*, 1(4):311–335, 1991.

[182] Elizabeth Phillips, Scott Ososky, Janna Grove, and Florian Jentsch. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics*

Society Annual Meeting, volume 55, pages 1491–1495. SAGE Publications Sage CA: Los Angeles, CA, 2011.

[183] Lane Phillips, Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. Distance perception in NPR immersive virtual environments, revisited. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pages 11–14. ACM, 2009.

[184] A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a computer agent with a humanoid robot. In *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 145–152, March 2007.

[185] William L. Prosser. Privacy. In Ferdinand David Schoeman, editor, *Philosophical dimensions of privacy: An anthology*. Cambridge University Press, 1960.

[186] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, volume 3, page 5, 2009.

[187] James Rachels. Why privacy is important. *Philosophy & Public Affairs*, pages 323–333, 1975.

[188] N. Raval, A. Srivastava, K. Lebeck, L. P. Cox, and A. Machanavajjhala. MarkIt: Privacy Markers for Protecting Visual Secrets. In *In Proceedings of the Workshop on Usable Privacy and Security for Wearable and Domestic ubIquitous DEvices (UPSIDE)*, 2014.

[189] Joseph Reagle and Lorrie Faith Cranor. The platform for privacy preferences. *Communications of the ACM*, 42(2):48–55, 1999.

[190] Neil Richards and Woodrow Hartzog. Taking trust seriously in privacy law. *Stan. Tech. L. Rev.*, 19:431, 2015.

[191] Neil M. Richards and William D. Smart. How Should the Law Think About Robots? In *Proceedings of We Robot: The Inaugural Conference on Legal and Policy Issues Relating to Robotics*, Coral Gables, FL, 2012.

[192] John M. Roberts and Thomas Gregor. Privacy: A cultural view. In J. Roland Pennock and John W. Chapman, editors, *Privacy and Personality*. Transaction Publishers, 1971.

[193] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3):309–314, 2004.

[194] Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404, 1998.

[195] Matthew Rueben and William D Smart. Privacy in human-robot interaction: Survey and future work. In *We Robot 2016: the Fifth Annual Conference on Legal and Policy Issues relating to Robotics*. University of Miami School of Law, 2016. Discussant: Ashkan Soltani, Independent Researcher.

[196] Matthew Rueben, Cindy M. Grimm, Frank J. Bernieri, and William D. Smart. A taxonomy of privacy constructs for privacy-sensitive robotics. arXiv:1701.00841v1 [cs.CY].

[197] Matthew Rueben, Frank J Bernieri, Cindy M Grimm, and William D Smart. User feedback on physical marker interfaces for protecting visual privacy from mobile robots. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 507–508. IEEE Press, 2016. Late-Breaking Report.

[198] Matthew Rueben, Frank J Bernieri, Cindy M Grimm, and William D Smart. Evaluation of physical marker interfaces for protecting visual privacy from mobile robots. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO–MAN 2016)*, pages 787–794. IEEE, 2016.

[199] Matthew Rueben, Frank J Bernieri, Cindy M Grimm, and William D Smart. Framing effects on privacy concerns about a home telepresence robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 435–444. ACM, 2017.

[200] Matthew Rueben, Alexander Mois Aroyo, Christoph Lutz, Johannes Schmölz, Pieter Van Cleynenbreugel, Andrea Corti, Siddharth Agrawal, and William D. Smart. Themes and research directions in privacy-sensitive robotics. In *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. IEEE, 2018. In press.

[201] Stuart Russell and Peter Norvig. *Artificial intelligence: A modern approach*. Prentice Hall, 3 edition, 2009.

[202] Radu Bogdan Rusu, Zoltan Csaba Marton, Nico Blodow, Andreas Holzbach, and Michael Beetz. Model-based and learned semantic object labeling in 3d point cloud maps of kitchen environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3601–3608. IEEE, 2009.

[203] Fredrik Rydén and Howard Jay Chizeck. Forbidden-region virtual fixtures from streaming point clouds: Remotely touching and protecting a beating heart. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3308–3313. IEEE, 2012.

[204] Fredrik Rydén and Howard Jay Chizeck. A method for constraint-based six degree-of-freedom haptic interaction with streaming point clouds. In *2013 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2353–2359. IEEE, 2013.

[205] O. Fredrik Rydén. *Real-Time Haptic Interaction with Remote Environments using Non-contact Sensors*. PhD thesis, 2013.

[206] Selma Sabanovic, Marek P Michalowski, and Reid Simmons. Robots in the wild: Observing human-robot social interaction outside the lab. In *9th IEEE International Workshop on Advanced Motion Control*, pages 596–601. IEEE, 2006.

[207] Mukesh Saini, Pradeep K Atrey, Sharad Mehrotra, and Mohan Kankanhalli. W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications*, 68(1): 135–158, 2014.

[208] Satoru Satake, Hajime Iba, Takayuki Kanda, Michita Imai, and Yoichi Morales Saiki. May I talk about other shops here?: modeling territory and invasion in front of shops. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 487–494. ACM, 2014.

[209] Burkhard Schafer and Lilian Edwards. "I spy, with my little sensor": fair data handling practices for robots between privacy, copyright and security. *Connection Science*, 29(3):200–209, 2017.

[210] Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow. Experiences with sparky, a social robot. In *Socially Intelligent Agents*, pages 173–180. Springer, 2002.

[211] Paul Schermerhorn, Matthias Scheutz, and Charles R. Crowell. Robot Social Presence and Gender: Do Females View Robots Differently Than Males? In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, HRI '08, pages 263–270, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-017-3.

[212] Matthias Scheutz, Scott A DeLoach, and Julie A Adams. A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making*, 11(3):203–224, 2017.

[213] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Y. Goldberg. Respectful Cameras: Detecting Visual Markers in Real-Time to Address Privacy Concerns. In *Proceedings of the IEEE/RSJ International Conference on Robots and Systems (IROS)*, pages 971–978, San Diego, CA, 2007.

[214] Ferdinand David Schoeman. *Philosophical dimensions of privacy: An anthology*. Cambridge University Press, 1984.

[215] Trenton Schulz and Jo Herstad. Walking away from the robot: Negotiating privacy with a robot. In *Proceedings of British HCI Conference (BISL)*, pages 1–6, 2017.

[216] Rachel Sebba and Arza Churchman. Territories and territoriality in the home. *Environment and Behavior*, 15(2):191–210, 1983.

[217] Elaine Sedenberg, John Chuang, and Deirdre Mulligan. Designing commercial therapeutic robots for privacy preserving systems and ethical research practices within the home. *International Journal of Social Robotics*, 8(4): 575–587, 2016.

[218] Luis Sentis and Oussama Khatib. A Whole-Body Control Framework for Humanoids Operating in Human Environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2641–2648, 2006.

[219] Drew Simshaw, Nicolas Terry, Kris Hauser, and ML Cummings. Regulating healthcare robots: Maximizing opportunities while minimizing risks. *Rich. JL & Tech.*, 22:1, 2015.

[220] David Sirkin, Brian Mok, Stephen Yang, and Wendy Ju. Mechanical ottoman: how robotic furniture offers and withdraws support. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 11–18. ACM, 2015.

[221] William D Smart and Neil M Richards. How the law will think about robots (and why you should care). In *2014 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 50–55. IEEE, 2014.

[222] H Jeff Smith, Tamara Dinev, and Heng Xu. Information privacy research: an interdisciplinary review. *MIS quarterly*, 35(4):989–1016, 2011.

[223] Daniel J. Solove. *Understanding privacy*. Harvard University Press, Cambridge, MA, 2008.

[224] Daniel J Solove and Woodrow Hartzog. The FTC and the new common law of privacy. *Colum. L. Rev.*, 114:583, 2014.

[225] Robert Sommer. *Personal Space: The Behavioral Basis of Design*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1969.

[226] Robert Sommer and Franklin D Becker. Territorial defense and the good neighbor. *Journal of personality and social psychology*, 11(2):85, 1969.

[227] Marco Spadafora, Victor Chahuneau, Nikolas Martelaro, David Sirkin, and Wendy Ju. Designing the behavior of interactive objects. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 70–77. ACM, 2016.

[228] Kristen Stubbs, Debra Bernstein, Kevin Crowley, and Illah R Nourbakhsh. Long-term human-robot interaction: The personal exploration rover and museum docents. In *AIED*, pages 621–628, 2005.

[229] Suitable Technologies. *Beam Remote Presence System*. 2012. https://www.suitabletech.com/.

[230] Jian Sun, Lu Yuan, Jiaya Jia, and Heung-Yeung Shum. Image completion with structure propagation. In *ACM Transactions on Graphics (ToG)*, volume 24, pages 861–868. ACM, 2005.

[231] JaYoung Sung, Henrik I Christensen, and Rebecca E Grinter. Robots in the wild: understanding long-term use. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 45–52. ACM, 2009.

[232] Dag Sverre Syrdal, Michael L. Walters, Nuno Otero, Kheng Lee Koay, and Kerstin Dautenhahn. He knows when you are sleeping—privacy and the personal robot companion. In *Workshop on the Human Implications of Human-Robot Interaction, Association for the Advancement of Artificial Intelligence (AAAI '07)*, pages 28–33, 2007.

[233] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and Emergent Behaviour and Complex Systems*, 2009.

[234] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[235] Leila Takayama, Victoria Groom, and Clifford Nass. I'm Sorry, Dave: I'm Afraid I Won't Do That: Social Aspects of Human-agent Conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2099–2108, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1519021.

[236] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76. ACM, 2011.

[237] Deborah Tannen. What's in a frame? Surface evidence for underlying expectations. *Framing in discourse*, 14:56, 1993.

[238] Robert Templeman, Mohammed Korayem, David Crandall, and Apu Kapadia. Placeavoider: Steering first-person cameras away from sensitive spaces. In *Network and Distributed System Security Symposium (NDSS)*, 2014.

[239] The United Nations. The Universal Declaration of Human Rights, 1948.

[240] Kristen Thomasen. Liar, Liar, Pants on Fire! Examining the Constitutionality of Enhanced Robo-Interrogation. In *Proceedings of We Robot 2012*, University of Miami, 2012. Draft.

[241] Judith Jarvis Thomson. The Right to Privacy. *Philosophy & Public Affairs*, 4(4):295–314, July 1975. ISSN 0048-3915.

[242] Andrew J Tomarken. A psychometric perspective on psychophysiological measures. *Psychological Assessment*, 7(3):387, 1995.

[243] Meg Tonkin, Jonathan Vitale, Suman Ojha, Jesse Clark, Sammy Pfeiffer, William Judge, Xun Wang, and Mary-Anne Williams. Embodiment, privacy and social robots: May I remember you? In *International Conference on Social Robotics*, pages 506–515. Springer, 2017.

[244] John J. Trinckes, Jr. Section 2.6: Social Engineering and HIPAA. In *The Definitive Guide to Complying with the HIPAA/HITECH Privacy and Security Rules*, page 472. Auerbach Publications, 2012. ISBN 9781466507678.

[245] Karthikeyan Vaiapury, Akil Aksay, and Ebroul Izuierdo. GrabcutD: Improved Grabcut using Depth Information. In *Proceedings of the 2010 ACM Workshop on Surreal Media and Virtual Cloning (SMVC)*, pages 57–62, 2010.

[246] VGo Communications. *VGo Robotic Telepresence for Healthcare, Education, and Business*. 2012. http://www.vgocom.com/.

[247] Jonathan Vitale, Meg Tonkin, Sarita Herse, Suman Ojha, Jesse Clark, Mary-Anne Williams, Xun Wang, and William Judge. Be more transparent and users will like you: A robot privacy and user experience design experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 379–387. ACM, 2018.

[248] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. A survey of user-centered design practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 471–478. ACM, 2002.

[249] Alan R Wagner. An autonomous architecture that protects the right to privacy. In *Proceedings of AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. AAAI, 2018.

[250] Tedra A. Walden, Paul A. Nelson, and Dale E. Smith. Crowding, privacy, and coping. *Environment and Behavior*, 13(2):205–224, 1981.

[251] Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.

[252] Rolf H. Weber. Internet of Things—New security and privacy challenges. *Computer Law & Security Review*, 26(1):23–30, 2010.

[253] Karl E Weick. *Sensemaking in organizations*, volume 3. Sage, 1995.

[254] Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, 1993.

[255] Alan F. Westin. *Privacy and Freedom*. Athenaeum, New York, NY, 1967.

[256] Richmond Y Wong and Deirdre K Mulligan. These aren't the autonomous drones you're looking for: investigating privacy concerns through concept videos. *Journal of Human-Robot Interaction*, 5(3):26–54, 2016.

[257] Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control*, pages 750–755. IEEE, 2006.

[258] Sarah N. Woods, Michael L. Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 51–58. IEEE, 2006.

[259] Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. What does the robot think? transparency as a fundamental design requirement for intelligent systems. In *Proceedings of the IJCAI Workshop on Ethics for Artificial Intelligence*, New York, 2016.

[260] Stephen Yang, Brian Ka-Jun Mok, David Sirkin, Hillary Page Ive, Rohan Maheshwari, Kerstin Fischer, and Wendy Ju. Experiences developing socially acceptable interactions for a robotic trash barrel. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 277–284. IEEE, 2015.

[261] Sarah K. Zeegers, Christine A. Readdick, and Sally Hansen-Gandy. Day-care Children's Establishment of Territory to Experience Privacy. *Children's Environments*, 11(4):265–271, December 1994. ISSN 2051-0780.

[262] Chenyang Zhang, Yingli Tian, and Elizabeth Capezuti. Privacy preserving automatic fall detection for elderly using rgbd cameras. In *Proceedings of the 13th international conference on Computers Helping People with Special Needs*, pages 625–633. Springer-Verlag, 2012.

[263] Qiang Alex Zhao and John T. Stasko. Evaluating image filtering based techniques in media space applications. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 11–18. ACM, 1998.

APPENDIX

A Interview Guide for "Mobile Shoe Rack" Study

Here we include the interview guide used in the study presented in Chapter 6. See Section 6.6 for a description of the interview procedure and analysis.

General Prompts

- "Let's focus on the last time you came to class. When was that? Tell me what you remember about the mobile shoe rack from that day."
- (Follow-ups:)
  - "Tell me about that."
  - "Do you remember what you thought about that?" "Any other thoughts?"
  - "What did you notice?" "Notice anything else?"
  - "Did you learn anything from [that]?" "Anything else?"
  - "What stood out to you about the mobile shoe rack? What did you notice?"
- (Prompts to make sure they've covered all their interactions with the "Mobile Shoe Rack":)
  - "Did anything else happen with the mobile shoe rack that day?"

  - "Did you see any other people interact with the mobile shoe rack?"
  - "Have you had any conversations with other people about the mobile shoe rack?"

- (If this was their first experience with the shoe rack:)
  - "Do you remember what you were expecting before you first saw the shoe rack?" "How did it compare to your expectations?"

Moving beyond their experiences in the hallway . . .

- "What are your opinions about the mobile shoe rack? What do you think of it?"
- "What are some things you like about the mobile shoe rack? How about some things you don't like? Do you have any ideas for how to make it better?"
- "Describe to me what the mobile shoe rack does. Describe all its actions."
- "Pretend I'm a friend of yours who has never seen the mobile shoe rack and you want to describe it to me. What would you tell me?"
- "Would you want this mobile shoe rack in your home?" "Can you think of a place where the mobile shoe rack would be a really good idea?" "... how about where it would be a really *bad* idea?" "Why?"
- "Is there anything you'd like to know about the mobile shoe rack at this point? that you're wondering about? ... What else?"

If they've seen the MSR more than once:

- "You've been around the mobile shoe rack a few times now. Have you learned anything about it since the first time you saw it?"
- "...Is there anything you notice now that you didn't notice at first?"

At the final meeting, right before debriefing them, we can ask about our DVs more directly:

- "Give us your best guess about the mobile shoe rack's capabilities. What can it do?"
- (ONLY if they need more prompting should you use these prompts:)
  - "Does it record/store things? Like what?"
  - "Can it see? Where and how far? Hear? Where and how far? How do you know? Understand speech?"
  - "Can it distinguish humans from the furniture? Like if you put a trash can in front of it?"
  - "Can it tell when your shoes are on vs. off? If you're in the process of taking them off?"
  - "Does it know when you've interacted with it?"
  - "If you sat in front of it again, do you think it would know you already gave it your shoes?"

- "How much was it paying attention to things? Did it notice you? Other people? How much attention to detail?"
- "Do you think it remembered you from day to day?"
- "Whether you're late? Can it tell how you look in your workout clothes? Whether you were awkward or clumsy or slow? If not, did you feel awkward anyway?"
- "How did you figure that out? How long did it take you? Tell us the story."