

Social Robots in Service Contexts: Exploring the Rewards and Risks of Personalization and Re-embodiment

Samantha Reig
Carnegie Mellon University
Pittsburgh, PA, USA
sreig@cs.cmu.edu

Isabel Won
Johns Hopkins University
Baltimore, MD, USA
iwon1@jhu.edu

Michal Luria
Carnegie Mellon University
Pittsburgh, PA, USA
mluria@cs.cmu.edu

Aaron Steinfeld
Carnegie Mellon University
Pittsburgh, PA, USA
steinfeld@cmu.edu

John Zimmerman
Carnegie Mellon University
Pittsburgh, PA, USA
johnz@cs.cmu.edu

Elsa Forberger
University of Minnesota-Twin Cities
Minneapolis, MN, USA
forbe140@umn.edu

Jodi Forlizzi
Carnegie Mellon University
Pittsburgh, PA, USA
forlizzi@cs.cmu.edu

ABSTRACT

Social agents and robots are moving into front-line positions in brick and mortar services, taking on roles where they directly interact with customers. These agents could potentially recognize customers to personalize service. Will customers like this, or might they feel monitored and profiled? Robots could also re-embodiment (move their “personality” between one body and another) in order to take on multiple roles that are typically performed by different people. Will this make customers feel more taken care of, or will it raise concerns about the robot’s competence and expertise? Our work investigates when robots should and should not recognize customers and re-embodiment. Our online study used storyboards to present possible future interactions between robots and customers across several different service contexts. Our findings suggest that people generally accept robots identifying customers and taking on vastly different roles. However, in some contexts, these robot behaviors seem creepy and untrustworthy.

CCS CONCEPTS

• Human-centered computing → Empirical studies in HCI.

KEYWORDS

human-robot interaction, human-agent interaction, storyboards, re-embodiment, personalization, interaction design

ACM Reference Format:

Samantha Reig, Michal Luria, Elsa Forberger, Isabel Won, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2021. Social Robots in Service Contexts: Exploring the Rewards and Risks of Personalization and Re-embodiment. In *Designing Interactive Systems Conference 2021 (DIS ’21)*, June 28–July 2, 2021, Virtual Event, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3461778.3462036>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

DIS ’21, June 28–July 2, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8476-6/21/06.
<https://doi.org/10.1145/3461778.3462036>

2, 2021, Virtual Event, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3461778.3462036>

1 INTRODUCTION

Social robots are increasingly taking on front-line service positions at places like airports, supermarkets, doctor’s offices, and in customer servicing. Though research has been done to understand how social robots should behave in service settings (e.g., [11, 19, 34]), there is much to be learned in the way of how they should behave when providing service over repeated interactions, what type of novel and nonhuman behaviors they should exhibit, and how the acceptability of certain behaviors may or may not generalize across service contexts. Different behaviors may be more or less suitable for different service contexts. When should these robots act like people, and when should they leverage their unique robotic capabilities to do things people cannot do, or might struggle to do? Our work probes two behaviors across a broad range of service situations: (1) Should robots identify customers? (2) Should robotic agents re-embodiment and take on multiple service roles?

Human-computer interaction (HCI) frequently makes use of immersion and simulation to investigate possible futures. This reduces the risk of building technical systems no one wants to use. Wizard-of-Oz (WoZ)—likely the best-known method of simulating interactions with systems that do not yet exist—uses researchers in the loop. In WoZ, researchers take on responsibilities of imagined technical systems, which allows them to investigate specific questions about the systems and their behaviors. User Enactments are a specific type of WoZ, where researchers make use of props, actors, and simulated environments to immerse participants in a particular future [20]. UE studies pose a trade-off between the tight experimental control needed for hypothesis-driven work and the ability to more broadly explore areas where less is known about how technology should behave.

A recent UE study investigated re-embodiment social robots: robots that move their social presence from one robot body to another in order to follow and continually meet a customer’s unfolding needs [14]. This study drew on findings from service theory

showing that customers often wish to interact with a single front-line worker (e.g., talking to the same person for both service and sales) [24]. The study placed participants in different service scenarios in which robots re-embodied. Participants often thought that re-embodiment could make a service experience more convenient, seamless, and pleasantly personal. However, they raised concerns when re-embodiment involved a large change in expertise, such as when a robot in a hospital served both as a receptionist and as a clinical technician that took a patient's X-ray. A different recent UE study explored whether robots should reveal personal information about customers when interacting in public spaces such as lobbies or retail store aisles [26]. It found that participants felt uncomfortable being identified in interactions where they would not expect a human worker to know them.

Our research builds on these two earlier UE studies. We set out to better understand when social robots working as front-line service agents should act more like human service workers, and when they should take advantage of capabilities that would be difficult or impossible for people (e.g., instantaneously accessing customer records, enacting many different roles, or moving their "identity" from one body to another). We wanted to gain insight into how different service contexts might influence people's perceptions of robot behavior. In addition, since UE studies work with small numbers of participants (due to their complexity to set up and execute), we wanted to investigate whether the preferences observed from small samples might generalize by using a larger audience.

We conducted an online study using storyboards [33] that described common service encounters that many people would find familiar. However, our storyboards featured robots in roles that are typically populated by human service agents. Using familiar situations helps participants reflect on the future based on their felt-experience of the present. Our use of storyboards to communicate these familiar yet future experiences allowed us to explore the two selected concepts (customer identification and robot re-embodiment) across several service contexts in a single study. Distributing the storyboards via an online study recruitment platform enabled us to rapidly determine whether insights from UE studies conducted with small numbers of participants could scale, and whether there are differences in the appropriateness of robot behaviors across different service contexts.

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

(see [8, 16]) on a variety of topics can draw on this method to advance understanding.

2 RELATED WORK

Our work builds upon previous research on service theory, and on research involving socially interactive service agents and robots and agents that re-embodiment and co-embodiment.

2.1 Service research in retail contexts

Service encounters, or situations in which an employee engages with a customer, are an ongoing topic of service research (e.g., [13, 27]). Researchers have identified three encounter types that particularly improve a customer's satisfaction with a service: when employees (i) respond to a customer's expressed needs or a request, (ii) provide unprompted and unexpected service, and (iii) recover well from a service breakdown [1]. This concept of a service encounter is highly relevant for designing interactive service systems, and is of particular concern for our research, where the front-line employee's role is performed by a social robot.

Service researchers have characterized personalization in service contexts as a "double-edged sword" in that it adds value, but also increases risks by making customers feel tracked [35]. Nevertheless, people tend to perceive data collection as less creepy when they understand why it is being collected in the first place [23]. Service researchers also note a paradox with personalized, online interactions. Customers experience them as impersonal because they interact with a computer rather than a person. However, online service encounters provide rich opportunities for personalization that would be impossible for human service agents to deliver [28, 35]. Our work probes this issue by exploring how social robots might span this gap. They can provide customers with a sense of familiarity to make interactions feel more personal, and they can personalize their performance for customers similar to how online systems personalize information display.

Finally, service theory shows that customers desire a "single face" for their interactions with a service; they do not want to talk to different people about sales and service [24]. This presents a significant challenge to services: they cannot train all of the required expertise into a single employee. Our work considers this as a potential opportunity area for social robots working in service contexts. Unlike humans, robots can take on several areas of expertise.

2.2 Social behaviors for service robots

Prior work shows that people tend to make assumptions about a robot's competence and expertise based on its form [22] and gender [30]. When a robot is not perceived as competent for the job it is doing, it impacts trust and acceptance of the robot [3]. Prior work shows that for social robots working within services, the specific context influences people's perceptions of trust, preferences for personalization, and their overall experience [14]. Perceptions of robots are also impacted by social and organizational norms [18], which impact the appropriateness of specific robot behaviors (e.g., [10]).

When people interact with robots, they often disclose private information [25], especially with "expressive" social robots [17]. They also tend to underestimate the risks of having their shared data

collected and used [12]. Simultaneously, people express concerns with having their data collected and kept by robots in both public and private spaces. Recent work that looked into robot behavior in homes found that people are concerned with robots that “watch, listen, and record” [15]. Other work examined reactions to a robot in a coffee shop, and found that people felt uncomfortable with a robot that listened to their conversations and looked them up in databases [7]. Whether the data was used for positive or negative purposes also impacted acceptance of this behavior. Studies of online behavior show that when people benefit from the use of personal data, they are less likely to perceive that use as creepy [23]. To extend this point, we examine whether providing *value* enabled by data use may impact the acceptability of data collection.

Prior work has been done on robot re-embodiment (also called “agent migration”) in lab and field settings [9, 21]. Previous work suggested that people are comforted by a single re-embodiment agent that follows them from a private setting to a service setting [26]. One study found that re-embodiment increased perceptions of trustworthiness, competence, and likeability [31]. Our study tackles the question of how to design re-embodiment robots for a range of diverse services. We focus specifically on how different service contexts impact expectations and perceptions of robot behavior.

3 CLAIMS

We sought to dive more deeply into findings that were raised in the literature. In an iterative ideation process, we narrowed down to two prior aspects of personalized service interactions: user recognition and robot re-embodiment. We set out to better understand *claims* that we derived from the literature on both aspects, which suggested that people’s reactions are strongly influenced by the service context. We emphasize that our proposal and analysis of these claims is purposefully *exploratory*. Rather than assert a prediction and then use empirical methods to test it, we use the claims to guide an evolving understanding of intermediate-level knowledge (see [8, 16]) of design concepts.

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

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

With respect to robot *re-embodiment*:

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

4 RESEARCH APPROACH

Designing the behavior of social robots working within brick and mortar services requires many choices, and each decision likely impacts a customer’s holistic experience. Given this nearly unbounded space of investigation, design methods offer an effective approach for gaining insights. Human-computer interaction (HCI) research notes the tension between scientific research that seeks to use complex instruments to exert control over phenomena, and design work that gleans knowledge from complexity through the use of simple

tools [29]. Design methods allow researchers to rapidly explore a broad set of design choices and future situations. This type of exploration is less accessible with hypothesis-driven research, in which many aspects of a design need to be carefully controlled in order to generate new knowledge. We used the method of Speed Dating with storyboards [5]. The mid-level fidelity of *storyboards* allows researchers to rapidly iterate pilots to progress towards stimuli that effectively probe a study’s research goals. They also allow participants to experience small sips of many different situations, which helps them gain higher-level insights on what they actually want and expect [36]. This method is useful for drawing participants’ attention to specific examples of technology behaviors, especially ones that are set in the future or that do not exist yet [2, 15, 33].

4.1 Creating storyboards

Over several months, 7 researchers iteratively generated scenarios and then storyboards for a diverse set of service contexts. We structured our brainstorming around situations that were socially complex. Many HRI and storyboard studies focus on one-to-one interactions—but service interactions involve groups, interpersonal relationships, and public spaces. Similar to [14] and [26], we aimed to produce scenarios that were novel and clearly set in the future, yet realistic enough that participants could readily imagine them occurring in their lives. We initially brainstormed 16 scenarios. We then refined these through four rounds of piloting involving 310 participants. We narrowed down our set of claims based on how the storyboards were received. After each pilot, we retained the scenarios that had the most traction and were most conducive to exploring design nuances in different service settings. Through multiple rounds of piloting, we gleaned insight into which service settings were sparking the most reflection from participants, and which storyboard pairs generated the most telling comparisons and contrasts in the data. We relied on our collective judgment and drew on our evolving knowledge of the design space and research method as we downselected to the final set of storyboards. Piloting also enabled us to identify anything that was confusing (e.g., overly complex robot dialogue) or misinterpreted (e.g., interactions we expected to deliver value but where participants perceived no value). Piloting resulted in 11 final storyboards related to our claims: 4 that addressed C1-Pers, 3 that addressed C2-Pers, and 4 that addressed C3-Re.

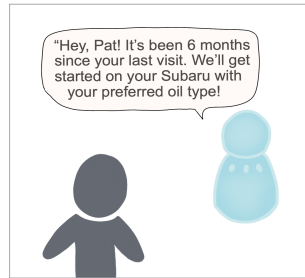
4.1.1 Final storyboards. We created two versions for each final storyboard; one capturing the assumed preference according to the claim, and one pushing against it. Figure 1 provides examples of storyboard pairs addressing C1-Pers and C3-Re. In the C1-Pers example pair, Pat, a customer, is recognized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

Personal
(expectation to be recognized)

Pat gets an oil change



Pat visits a local auto repair shop for an oil change.



The robotic agent that works at the shop recognizes and greets Pat.

Impersonal
(no expectation to be recognized)

Pat goes to a carnival



Pat is visiting a carnival with some friends.



Pat is recognized and greeted by a robotic agent upon entering.

↑ [C1-PERS] ↑

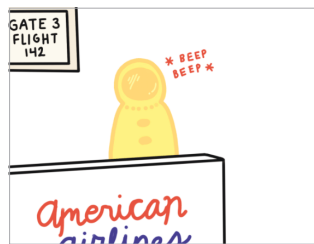
↓ [C3-RE] ↓

Re-embodiment from a gate agent to an in-flight entertainment system

Jerry flies home for Thanksgiving



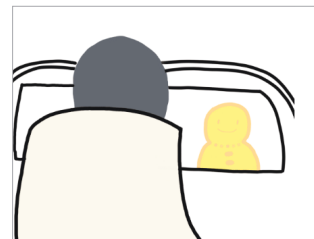
It's Thanksgiving weekend and Jerry is flying home to visit his family. Jerry hands his ticket to a boarding agent robot, who introduces itself as Bob and wishes Jerry a good flight.



After all the passengers have been checked in, Bob shuts off to transfer his presence.



Shortly after, Jerry hears another agent start to speak over the intercom, introducing itself as the co-pilot of this flight.



Later, boarding agent Bob re-appears on Jerry's screen, this time as Jerry's personal in-flight entertainment guide, and tells him that his connecting flight has a 30 minute delay.

Re-embodiment from a gate agent to an airplane co-pilot

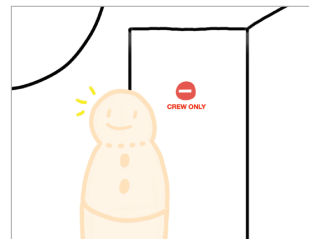
Jerry flies home for Thanksgiving



It's Thanksgiving weekend and Jerry is flying home to visit his family. Jerry hands his ticket to a boarding agent robot, who introduces itself as Bob and wishes Jerry a good flight.



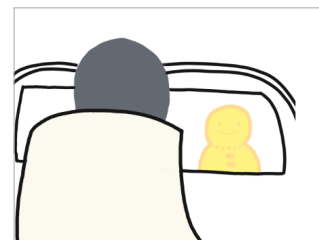
After all the passengers have been checked in, Bob shuts off to transfer his presence.



A larger robot near the cockpit lights up, and Bob re-appears on its screen. Bob heads inside the cockpit.



Shortly after, Jerry hears Bob's voice over the plane intercom re-introducing himself as the co-pilot of this flight.



Bob re-appears on Jerry's screen, telling Jerry that his connecting flight has a 30 minute delay.

Figure 1: An example of two storyboard pairs that were compared to address C1-Pers and C3-Re. Top left: Pat is identified by a customer service robot at an auto shop. Top right: Pat is identified by a robotic employee at a carnival. Bottom left: Jerry interacts with an airline gate agent robot that re-embodies to function as the plane's in-flight entertainment system. Bottom right: Jerry interacts with an airline gate agent robot that re-embodies to function as the plane's co-pilot.

We also purposely kept the narratives fairly straightforward and simple. Research on the fidelity of prototypes (e.g., [4, 5, 33]) details the need to choose the “right fidelity” to keep a focus on the phenomena of concern and has shown that storyboards allow for more breadth of search in exploratory research. As stimuli become more specific, the question changes from “Would people want a robot to exhibit this behavior?” to “Would people want *this* robot to exhibit this behavior?” Our storyboards were simple to strike a balance: examples had to be specific enough to give participants something to reflect on, but not so specified that they would overly constrain interpretations and reflections.

Customers were depicted in a consistent, 2D visual style that de-emphasized gender and racial cues, allowing participants to more easily envision themselves in the depicted situation. We drew robots of different forms, sizes, and colors for different tasks and contexts. We also used gender neutral names for robots to reduce any gender effects, as seen in prior work [30]. See the Supplementary Material for the full set of storyboards and questionnaires.

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

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

C3-Re: The third set centered on robot re-embodiment and differing expertise. Prior work [14] found that people are likely to be concerned about a re-embodiment robot having the expertise required to do different jobs. In this work, we used specific examples of contexts requiring different kinds of expertise to examine *why* this might be the case. We probed at whether the “social status” or “prestige” that people associate with different jobs might play into these concerns. We also explored whether the similarity of the roles assumed by a re-embodiment robot might impact concerns about expertise. In these storyboard pairs, we varied the similarity of the roles that the robot played before the re-embodiment and after the re-embodiment. The storyboard settings (each with two different versions of the “target” of the re-embodiment) were: 1) a hotel, 2) air travel, 3) physical therapy and massage, and 4) a dentist’s office.

4.2 Participants

We recruited 204 participants through the online survey research platform Prolific. To be included, participants had to (1) be at least

18 years of age, (2) be fluent in English, (3) have previously completed at least 50 submissions on Prolific, and (4) have at least a 90% approval rate for previous submissions. Participants were paid \$5.00 USD each. Our study employed attention checks, and 197 participants answered these questions correctly. Most participants (107, 54%) were between the ages of 18-30. Sixty-nine (35%) were 30-45, and twenty (10%) were 45 or older.¹ One did not report their age. Ninety-nine (50%) identified as female, 94 (48%) identified as male, 1 identified as non-binary, 2 self-described as genders not represented in our multiple choice options, and 1 chose not to disclose their gender. The study was approved by our university’s Institutional Review Board.

4.3 Procedure

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

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

4.4 Measures

For storyboards that focused on C1 (personal and impersonal contexts) and C2 (added value), we assessed perceptions of the robot as *creepy* and perceptions of the encounter as *friendly*². For the storyboards that focused on C3 (re-embodiment), we assessed perceptions of the level of *prestige* and *expertise* of each role performed by the robot in the story, perceptions of the robot’s *competence*, belief that the service did a good job of creating an agent capable of *multitasking*, whether or not it was appropriate for the agent to *serve in both roles*, and whether or not the agent *should take on multiple responsibilities*. For all scenarios, we asked about perceptions of the encounter as an *improvement over the typical service experience*. Finally, for each scenario, an open-ended question asked participants to explain their ratings for the main variable of interest.

¹Our response categories forced any participant who was 30 years old or 45 years old to choose between two overlapping descriptions. Unfortunately, this error was not caught until after data collection.

²“Creepy” and “friendly” are not opposites, but they are concepts with opposing sentiments that could each possibly describe the way it feels to be spoken to in a very personal manner, especially by a robot. Our choice to use these particular positive-sentiment and negative-sentiment words was inspired by the themes from [26]. Perceptions of creepiness and friendliness were measured via two separate questions; we do not assume them to be mutually exclusive.

For the C1 and C2 storyboards, this question pertained to ratings for how *creepy* the agent's behavior was. For the C3 storyboards, it pertained to ratings of whether the agent should or *should not take on multiple responsibilities*. Because this is exploratory design research intended to garner direct feedback from participants, we designed the questionnaire items to directly ask about the concepts and perceptions we were interested to understand. Each question was analyzed as a single item. Figure 2 shows the response distributions for these variables.

4.5 Analysis

To analyze the qualitative data, we took a two-pronged approach. We first looked at the short-answer explanations associated with ratings at the extremes of the Likert items to check for extreme response bias [6]. Our review of this subset of the data suggested that people did indeed have strong positive and negative reactions to the scenarios. Two members of the research team then went back through the full data set, manually reading each response and annotating the findings. In doing so, they made note of which responses were associated with which storyboards, and where the participant's corresponding scale ratings fell (e.g., an individual response might be annotated with "superstore-no value, perceived creepiness=very high").³ This allowed us to interpret patterns in light of the comparisons we intended to draw between the different storyboard versions. Multiple research team members reviewed these notes and discussed key themes and insights in the data, leading to the insights we discuss in the Findings section and our design recommendations.

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

5 FINDINGS

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

5.1 Impersonal and personal settings

Differences in participant responses revealed nuances regarding the appropriateness of recognizing users in different service contexts. Prior work [26] found that people did not like service robots to recognize them in settings they viewed as impersonal (e.g., a department store), but desired it in settings where they expected personalized service (e.g., a doctor's office). This draws a distinction between two types of contexts: those in which the professional relationship between the service providers and the customer involves a degree of more-intimate interaction, and those in which it maintains more distance. Our findings suggest that **expectation to be (and appropriateness of being) identified by robots is a**

complex issue that cannot be reduced to a simple binary. Overall, being identified by a robot was perceived as less creepy in settings where customers expected to interact with the same service agent, and expected service to be personalized to their individual needs.

Across all C1-Pers scenarios, participants who found the agent to be creepy generally expressed one of three concerns. One, they opposed facial recognition. Many (n=20) commented explicitly on their **discomfort with facial recognition software being employed in a service setting**. None of the storyboards detailed how the robot recognized the customer—rather, participants inferred that it was using facial recognition (in fact, one participant, P29, commented that the robot was *not* creepy because they assumed it to *not* be using facial recognition). Two, participants did not want to be profiled. They explained that the robot's verbal disclosure of the amount of time since the customer's last visit showed the service collected an **unnecessary and uncomfortable amount of information**. (Participants who mentioned this included several of those who explicitly pointed to facial recognition as a concern and several others.) A robot's intention to be friendly or helpful was not enough to justify the profiling. Instead, participants came to their own conclusions about whether the service had reason enough to collect and use personal data. Three, participants shared that being identified by the robot was creepy because it **lacked the human-like characteristics** that could make this kind of interaction seem empathetic. This reaction was similar to findings in [26].

Participants who found the personal identification (which happened in all C1-Pers storyboards) less creepy commonly mentioned that it was a **friendly behavior that would make them feel welcome** (e.g., "I think it makes a more personalized experience and makes people feel more welcomed and seen,"—P106), or that it added value to the service experience (e.g., "I love being welcomed. A little compliment goes a long way for me,"—P140). Several participants thought that the robot's behavior was perfectly acceptable because it was no different from how a human in that position would behave (9 responses reflected this sentiment directly, and several others alluded to it). Interestingly, a couple of participants believed that a robot could do the human-facing, empathy-requiring aspects of the job *better* than a human could, and therefore, would prefer robots in service roles involving personal identification (P17 said that robots would not "disrespect and judge" as humans do).

5.1.1 Out-of-place identification. The carnival/auto shop storyboard pairing demonstrated the largest differential in perceptions of the agent as creepy. Perceived creepiness was higher in the carnival than in the auto shop, the auto shop was perceived to be a better improvement over the typical service experience, and the auto shop encounter was perceived as more friendly (see Table 1).

We infer that the large difference between the carnival and auto shop storyboards stems from the fact that **a carnival is at the extreme low end of the expectation to be identified spectrum**. For most, a carnival is a novelty event that is not available year-round. While some ride operators and ticket salespeople may be locals hired for a single gig, long-term employees or robotic service workers would likely travel. To be identified by them would be an anomaly. Here, when participants thought that the encounter was *not* creepy, they noted that it was "wholesome and harmless"

³We make statements about "participants who found the robot to be creepy" and similar generalizations throughout the Findings section. Such classification of participants is based on their scale ratings; i.e., whether their creepiness scores were lower or higher than zero.

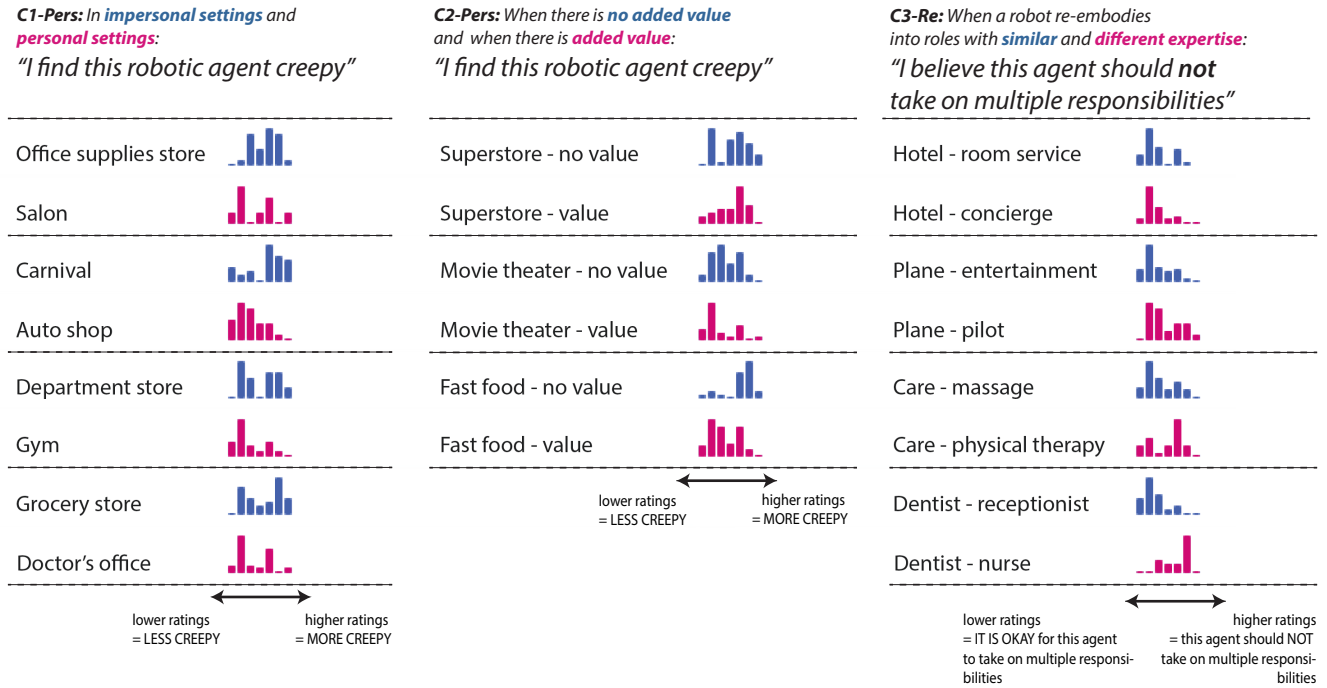


Figure 2: The distribution of ratings of creepiness (C1 and C2) and the belief that a robot should not re-embodiment into roles with different expertise (C3) for each of the 22 storyboards.

Table 1: M (SD) for each storyboard in the *Impersonal (I)* vs. *Personal (P)* set (C1-Pers). Ratings were on a scale from -3 (strong disagreement) to 3 (strong agreement). Values in bold with * are < .05. The higher value of each measure for each storyboard pair is in bold, blue text.

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

(P13), and that learning the wait time for a favorite ride was useful. When they were bothered by it, they **commented on the identification being out-of-place and unnecessary**: P45 said, “It would be strange to be recognized personally at an amusement park after a year and for them to know your favorite ride.” At $M = 0.5$, carnival creepiness was rated higher than any other storyboard (second-highest was the grocery store, $M = 0.3$). In contrast, participants commented that **in an auto shop, having data on a customer and their visit history is directly related to the service being**

provided (e.g., “I’m assuming the robotic agent just has documentation and a log of all its customers and their past services”—P130).

5.1.2 Identification that is necessary for the service. The doctor’s office was at the opposite end of the spectrum. In medical settings, correct and reliable identification is mission-critical. Participants’ explanations of their creepiness ratings spoke to this: P22 said, “Doctor offices keep files [...] it is not top secret information. It is reaffirming your identity and confirming,” and P175 said, “At the doctor’s office I am more comfortable with an AI that is aware of

my visit frequency.” A hair salon and gym are both somewhere in the middle. In these settings, it is fairly common to be recognized by workers, and in certain cases—for example, when getting a haircut from one’s favorite stylist, or when working with a personal trainer—not being identified correctly could be cause for alarm and/or be a detrimental experience. In these two contexts, participants who did not find identification creepy said that it was a useful feature (e.g., “It can help save time”, said P44 about the hair salon), that it made sense given the setting (e.g., “Their job is to remember things like this for the experience,” said P73 about the gym), and that it did not overstep an interpersonal boundary (e.g., “Not threatening in any way,” said P34 about the hair salon). This distinction bore out in the Likert ratings as well: creepiness was higher in the grocery store than in the doctor’s office, and higher in the department store than in the gym (see Table 1).

5.2 When identification adds value

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

5.2.1 Overt tracking. In each of the three storyboard pairs, the pattern was the same: First, **the no value storyboard was creepier** than the *value* storyboard. Second, the *value* story was a larger improvement over the typical service experience, and the robot’s behavior was more appropriate. Finally, people were more suspicious of robots in *no value* storyboards. However, not all of these differences were statistically significant (see Table 2).

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

5.2.2 Social privacy violations. A theme that emerged in C1 and C2 storyboards was concern about the possibility of social privacy violations. Participants were worried that in public settings, regardless of recognition by a robot or value of the service, **strangers might overhear conversations between a customer and the robot, which could affect their trust in the robot and in the**

service. This sentiment was particularly strong in the superstore scenario, where the robot said aloud that the customer was probably looking for diapers. Participants did not like the idea of this private information being made known to anyone within earshot, and thought it could even “threaten a customer’s safety” (P187). There were no significant differences in any of the ratings of the service encounter as creepy, better than the typical service experience, appropriate, or suspicious between the *value* and *no value* versions of this storyboard. It is likely that the public announcement of private information (which occurred in both storyboards) was so noticeable and so unappealing to many participants that it undermined their likelihood of caring about or even noticing the value-related difference between the two.

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

5.3 Re-embodiment and expertise

5.3.1 Multiple roles: why not? In general, **participants seemed comfortable with the idea of robots taking on multiple roles requiring different kinds of expertise**. In all eight of the C3 storyboards, average ratings for the statement “the agent should *not* take on multiple responsibilities” were relatively low (see Table 3). Justifications of these ratings evoked the convenience of streamlining the interaction (i.e., having one party keep track of all of the information across multiple touchpoints), good faith in a robot’s ability to handle multiple tasks at once better than a human, and a sense of indifference: “Why not?” (P41, P56, P72, P96, and others).

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

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

Table 2: M (SD) for each storyboard in the *Added Value* set (C2-Pers). Values in bold with * are < .05. The higher value of each measure for each storyboard pair is in bold, blue text.

Storyboard version	Value or no value	Creepy	Better than typical experience	Acts appropriately	Suspicious
SUPERSTORE	No value	0.18 (1.78)	0.16 (1.57)	0.42 (1.75)	-0.26 (1.83)
	Value	0.08 (1.75)	0.76 (1.44)	0.76 (1.63)	-0.43 (1.93)
	<i>p</i> -value	0.782	0.052	0.326	0.656
MOVIE THEATER	No value	-0.41 (1.61)	0.16 (1.36)	1.10 (1.42)	-0.94 (1.59)
	Value	-1 (1.73)	1.38 (1.65)	1.66 (1.39)	-0.94 (1.90)
	<i>p</i> -value	0.080	0.0001*	0.051	0.997
FAST FOOD	No value	0.61 (1.86)	-0.72 (1.54)	0.15 (1.51)	0.24 (1.75)
	Value	-0.52 (1.65)	1.10 (1.40)	1.48 (1.23)	-0.77 (1.88)
	<i>p</i> -value	0.002*	<.0001*	<.0001*	0.007*

Table 3: M (SD) for each storyboard in the *Re-embodiment and Expertise* set (C3-Re). Values in bold with * are < .05. The higher value of each measure for each storyboard pair is in bold, blue text.

Context	Second role	Better than typical experience	Should NOT take on multiple responsibilities	Should NOT serve in both of these roles	Competent
HOTEL	Food	0.84 (1.23)	-1.06 (1.52)	-0.94 (1.64)	1.92 (0.81)
	Concierge	0.92 (1.19)	-1.14 (1.40)	-1.16 (1.36)	1.68 (1.06)
	<i>p</i> -value	0.733	0.789	0.467	0.211
AIR TRAVEL	Seat-back	0.62 (1.24)	-0.88 (1.67)	-1.10 (1.61)	1.54 (1.18)
	Pilot	-0.04 (1.38)	-0.22 (1.79)	0.00 (1.90)	1.45 (1.28)
	<i>p</i> -value	0.014*	0.063	0.003*	0.714
PERSONAL CARE	Massage	0.50 (1.57)	-0.69 (1.71)	-0.62 (1.83)	0.98 (1.37)
	Physical therapy	-0.09 (1.85)	-0.09 (1.85)	0.83 (1.34)	0.02 (1.67)
	<i>p</i> -value	0.096	0.097	<.0001*	0.003*
DENTIST	Receptionist	0.53 (1.43)	-1.35 (1.42)	-1.22 (1.52)	1.74 (1.13)
	Receptionist + nurse	-0.50 (1.39)	0.46 (1.78)	N/A	1.64 (1.19)
	<i>p</i> -value	0.0004*	<.0001*	N/A	0.686

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

lacked qualifying comments that implied doubt about the ability to do multiple jobs.

Of the 17 participants who believed the agent embodying a boarding agent and a co-pilot should *not* take on multiple responsibilities (gave ratings to the right of zero), 10 called attention to the large difference between the two jobs. Participants explicitly called out **worries about risk and physical safety** as causes for concern: P22 said, “If they can’t focus on their job and get mixed up, that could be disastrous,” and P75 said, “More opportunity for something to go wrong. This especially applies to important responsibilities like piloting an airplane where there could be loss of life if something were to go wrong.” The *improvement over the typical service experience* ratings were higher in the entertainment system version than in the pilot version. Participants also had more concern with the robot serving in both roles in the pilot scenario (see Table 3). Both scenarios introduced a pilot robot (see Figure 1), which suggests that the *re-embodiment* aspect was what raised concern and discomfort.

5.3.4 Dentist’s office: empathy and training. In the other three contexts for C3, one robot only ever took on a maximum of two roles in a single storyboard. In the dentist storyboards, the robot either took on two roles (parking assistant, then receptionist) or

three roles (parking assistant, then receptionist, then dental nurse). Overall, the qualitative responses to this scenario looked similar to those from the air travel scenario. In general, people did not take issue with the same robot performing multiple roles. Those who did had concerns about expertise (e.g., P83 said a dental nurse was “a more specialized job”) and risk of unexpected events (e.g., P38 said, “The job of dental nurse should be done by a human so that they can monitor pain or anomalies when cleaning.”) As in the air travel scenario, a few participants expressed strong distaste without specific cause. The version in which the robot did *not* re-embodiment into the dental nurse was perceived as a significantly larger improvement over the typical service experience. People were also less concerned about the robot taking on multiple responsibilities when it did not serve in the role of the dental nurse (see Table 3).

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

Here, participants thought the service did a better job of creating a multitasking agent, believed the agent to be more competent, and were more comfortable with the robot serving in both roles in the masseuse scenario than in the physical therapy scenario (see Table 3). These findings suggest that people feel uncomfortable with re-embodiment when (1) one of the roles is high-risk, or (2) when the two roles are vastly different from each other and are taken from different service domains.

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

6 DISCUSSION

Our study revealed novel and critical insights about the way robots should and should not behave in service contexts. By situating three

concepts from prior work in several different contexts, we were able to draw comparisons across different service settings. Our study was motivated by knowledge that context matters when robots identify people and re-embodiment. In this study, we gained specific knowledge of *how* context matters when robots identify people and re-embodiment. We organize our discussion around the three claims, and propose specific design recommendations for each (Table 4). In addition, we offer reflection on our use of an online storyboard study as one way of advancing knowledge from UE studies.

6.1 Claims

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

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

C3-Re: *People will not want robots to re-embodiment when it involves a large change in social role or expertise.* The scenarios involving re-embodiment demonstrated concerns about multitasking, consistent with previous work [14]. Participants in favor of re-embodiment noted the value of multitasking in providing familiarity and a seamless experience for the customer, and increased efficiency for the service. They recognized that robots could theoretically multitask

Table 4: Implications for designing identifying and re-embodiment robots for different service settings.

Design Topic	Service Setting	Finding	Design Implication
Identification by a robot	Personal appointment services (e.g., hair salon, car service)	Being identified by a robot in a personal appointment setting can improve the service experience.	<i>If personalization is required for the service, a robot should identify the user. Otherwise, service robots should avoid identifying people.</i>
	Non-personal appointment services (e.g., department store, carnival)	Being identified by a robot in a non-personal appointment setting is perceived as creepy.	
	Personal and non-personal appointment services	Being identified by a robot that appears to use facial recognition is especially uncomfortable.	
Re-embodiment robots	Roles that are perceived as managing high risk situations (e.g., airplane pilot)	Robots that re-embodiment in high-risk situations are perceived negatively and as unsafe.	<i>In low-risk service contexts, robots can re-embodiment to provide a better service experience.</i>
	Roles that are perceived as managing low risk situations (e.g., hotel concierge)	People are accepting of robots that re-embodiment in low-risk service contexts.	
	Several tasks within one domain (e.g., room service, hotel maintenance, and concierge)	Robots that re-embodiment for different tasks that are clearly in the same domain are perceived more positively.	<i>If robots fulfill several tasks in one larger domain, re-embodiment can improve the service experience. A robot should not re-embodiment to do tasks in different domains.</i>
	Several tasks in different domains (e.g., haircut and massage therapy)	Robots that re-embodiment for tasks across different domains are perceived more negatively.	

better than humans, and believed that since they *can* take on multiple roles, they *should* take on multiple roles. However, they were concerned about multitasking robots when the jobs required very different skill sets.

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

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

6.2 Future directions

Our findings reveal directions for future work. In the C1 and C2 storyboards (about robots identifying people), we looked at a total of eleven specific contexts in which service robots might use personal information about customers in publicly observable interactions.

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

Through the C3 storyboards (on re-embodiment into robots with different roles and expertise), we began to explore possible definitions and impacts of expertise, role, status, and prestige for robots and AIs. Because artificial agents are not limited in the same ways as people are (e.g., they can exist in multiple places at once, they can have perfect memory, and they do not require as much time as humans do to “learn” how to demonstrate skill proficiency), these concepts will likely have different meanings for robots than they do for people. They undoubtedly will shape people’s impressions, comfort, and trust differently when exhibited by robots. The human-robot interaction and service design communities could

benefit from a deeper theoretical understanding of these social constructs as they apply to service robots, and we recommend that future work interrogate this. Finally, in the future, HRI researchers might consider exposing participants to in-person experiences derived from the scenarios in our storyboards via User Enactments and Wizard-of-Oz methods. A more personalized, higher-fidelity experience may reveal additional new insights on roles and expertise.

6.3 Reflection on methodology

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

6.4 Limitations

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

the service again, and individual and cultural differences (e.g., individualist vs. collectivist orientation [32]) are additional promising directions for future research.

7 CONCLUSION

In this work, we built on findings from low-fidelity studies on behavior designs for service robots. Our goal was to deepen our knowledge through a more structured, mid-fidelity study: We tested several storyboards that each addressed a single claim in a particular service situation. Our findings inspire design implications for creating appropriate robot identification and re-embodiment behaviors based on the service setting (see Table 4). Finally, this work contributes an example of how knowledge from initial exploratory research can be advanced.

ACKNOWLEDGMENTS

We thank our participants and the undergraduates in our lab who helped with piloting. This work was supported by the National Science Foundation (SES-1734456) and the National Aeronautics and Space Administration (80NSSC19K1133).

REFERENCES

- [1] Mary Jo Bitner, Bernard H. Booms, and Mary Stanfield Tetreault. 1990. The service encounter: diagnosing favorable and unfavorable incidents. *Journal of marketing* 54, 1 (1990), 71–84. Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- [2] Stacy M. Branham, Shahtab Wahid, and D. Scott McCrickard. 2007. Channeling creativity: Using storyboards and claims to encourage collaborative design. In *Workshop on Tools in Support of Creative Collaboration (part of Creativity & Cognition 2007)*, 1–4.
- [3] De'Aira Bryant, Jason Borenstein, and Ayanna Howard. 2020. Why Should We Gender? The Effect of Robot Gendering and Occupational Stereotypes on Human Trust and Perceived Competency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 13–21. <https://doi.org/10.1145/3319502.3374778>
- [4] Marion Buchenau and Jane Fulton Suri. 2000. Experience prototyping. In *Proceedings of the 3rd conference on Designing Interactive Systems: processes, practices, methods, and techniques (DIS '00)*. Association for Computing Machinery, New York, NY, USA, 424–433. <https://doi.org/10.1145/347642.347802>
- [5] Scott Davidoff, Min Kyung Lee, Anind K. Dey, and John Zimmerman. 2007. Rapidly Exploring Application Design through Speed Dating. In *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp '07)*. Springer-Verlag, Berlin, Heidelberg, 429–446.
- [6] Eric A. Greenleaf. 1992. Measuring extreme response style. *Public Opinion Quarterly* 56, 3 (1992), 328–351. Publisher: Oxford University Press.
- [7] Samarendra Hedao, Akim Williams, Chinmay Wadgaonkar, and Heather Knight. 2019. A Robot Barista Comments on its Clients: Social Attitudes Toward Robot Data Use. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 66–74. <https://doi.org/10.1109/HRI.2019.8673021>
- [8] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Trans. Comput.-Hum. Interact.* 19, 3, Article 23 (Oct. 2012), 18 pages. <https://doi.org/10.1145/2362364.2362371>
- [9] Chyon Hae Kim, Yumiko Yamazaki, Shunsuke Nagahama, and Shigeki Sugano. 2013. Recognition for psychological boundary of robot. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 161–162. <https://doi.org/10.1109/HRI.2013.6483551>
- [10] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: a longitudinal field experiment. *ACM Press*, 319. <https://doi.org/10.1145/2157689.2157804>
- [11] Min Kyung Lee, Jodi Forlizzi, Paul E. Rybski, Frederick Crabbe, Wayne Chung, Josh Finkle, Eric Glaser, and Sara Kiesler. 2009. The Snackbot: Documenting the Design of a Robot for Long-Term Human-Robot Interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (HRI '09)*. Association for Computing Machinery, New York, NY, USA, 7–14. <https://doi.org/10.1145/1514095.1514100>
- [12] Min Kyung Lee, Karen P. Tang, Jodi Forlizzi, and Sara Kiesler. 2011. Understanding Users' Perception of Privacy in Human-Robot Interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI '11)*. Association

- for Computing Machinery, New York, NY, USA, 181–182. <https://doi.org/10.1145/1957656.1957721>
- [13] Seung-Hee Lee. 2011. Managing Nongranting of Customers' Requests in Commercial Service Encounters. *Research on Language & Social Interaction* 44, 2 (April 2011), 109–134. <https://doi.org/10.1080/08351813.2011.567091>
- [14] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of Social Presence for Robots and Conversational Agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 633–644. <https://doi.org/10.1145/3322276.3322340>
- [15] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. <https://doi.org/10.1145/3313831.3376311>
- [16] Jonas Löwgren. 2013. Annotated portfolios and other forms of intermediate-level knowledge. *Interactions* 20, 1 (Jan. 2013), 30–34. <https://doi.org/10.1145/2405716.2405725>
- [17] Nikolas Martelaro, Victoria C. Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more: Designing HRI to encourage more trust, disclosure, and companionship. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 577–577. <https://doi.org/10.1109/HRI.2016.7451864>
- [18] Bilge Mutlu and Jodi Forlizzi. 2008. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. ACM Press, 287. <https://doi.org/10.1145/1349822.1349860>
- [19] Marketta Niemelä, Päivi Heikkilä, and Hanna Lammi. 2017. A social service robot in a shopping mall: expectations of the management, retailers and consumers. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 227–228.
- [20] William Odom, John Zimmerman, Scott Davidoff, Jodi Forlizzi, Anind K. Dey, and Min Kyung Lee. 2012. A fieldwork of the future with user enactments. In *Proceedings of the 2012 Conference on Designing Interactive Systems*. ACM Press, Newcastle Upon Tyne, United Kingdom, 338. <https://doi.org/10.1145/2317956.2318008>
- [21] Kohei Ogawa and Tetsuo Ono. 2008. ITACO: Effects to interactions by relationships between humans and artifacts. In *International Workshop on Intelligent Virtual Agents*. Springer, 296–307.
- [22] Steffi Paepcke and Leila Takayama. 2010. Judging a bot by its cover: An experiment on expectation setting for personal robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 45–52. <https://doi.org/10.1109/HRI.2010.5453268>
- [23] Chanda Phelan, Cliff Lampe, and Paul Resnick. 2016. It's Creepy, But It Doesn't Bother Me. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5240–5251. <https://doi.org/10.1145/2858036.2858381>
- [24] Adam A. Rapp, Daniel G. Bachrach, Karen E. Flaherty, Douglas E. Hughes, Arun Sharma, and Clay M. Voorhees. 2017. The role of the sales-service interface and ambidexterity in the evolving organization: A multilevel research agenda. *Journal of Service Research* 20, 1 (2017), 59–75. Publisher: Sage Publications Sage CA: Los Angeles, CA.
- [25] Alexander Reben and Joseph Paradiso. 2011. A Mobile Interactive Robot for Gathering Structured Social Video. In *Proceedings of the 19th ACM International Conference on Multimedia (MM '11)*. Association for Computing Machinery, New York, NY, USA, 917–920. <https://doi.org/10.1145/2072298.2071902>
- [26] Samantha Reig, Michal Luria, Janet Z. Wang, Danielle Oltman, Elizabeth Jeanne Carter, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2020. Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, Cambridge, United Kingdom, 289–297. <https://doi.org/10.1145/3319502.3374795>
- [27] Virpi Roto, Jung-Joo Lee, Tuuli Mattelmäki, and John Zimmerman. 2018. Experience Design meets Service Design: Method Clash or Marriage?. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–6. <https://doi.org/10.1145/3170427.3170626>
- [28] Roland T. Rust and Tuck Siong Chung. 2006. Marketing models of service and relationships. *Marketing science* 25, 6 (2006), 560–580. Publisher: INFORMS.
- [29] Erik Stolterman. 2008. The nature of design practice and implications for interaction design research. *International Journal of Design* 2, 1 (2008). Publisher: Chinese Institute of Design.
- [30] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior* 38 (2014), 75–84. Publisher: Elsevier.
- [31] Ravi Tejwani, Felipe Moreno, Sooyeon Jeong, Hae Won Park, and Cynthia Breazeal. 2020. Migratable AI. *arXiv:2007.05801 [cs]* (July 2020). <http://arxiv.org/abs/2007.05801> arXiv: 2007.05801.
- [32] Harry C. Triandis. 2001. Individualism-collectivism and personality. *Journal of personality* 69, 6 (2001), 907–924. Publisher: Wiley Online Library.
- [33] Khai N. Truong, Gillian R. Hayes, and Gregory D. Abowd. 2006. Storyboarding: An Empirical Determination of Best Practices and Effective Guidelines. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 12–21. <https://doi.org/10.1145/1142405.1142410>
- [34] Astrid Weiss, Markus Bader, Markus Vincze, Gert Hasenhütl, and Stefan Moritsch. 2014. Designing a Service Robot for Public Space: An "Action and Experiences" - Approach. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI '14)*. Association for Computing Machinery, New York, NY, USA, 318–319. <https://doi.org/10.1145/2559636.2559791>
- [35] Xiaoni Zhang and V.R. Prybutok. 2005. A consumer perspective of E-service quality. *IEEE Transactions on Engineering Management* 52, 4 (2005), 461–477. <https://doi.org/10.1109/TEM.2005.856568>
- [36] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (March 2017), 30–50. <https://doi.org/10.1016/j.sheji.2017.08.003>