Impacts of Robot Learning on User Attitude and Behavior

Anonymous

ABSTRACT

With an aging population and growing shortage of nurses and caregivers, the need for in-home robots is increasing. However, it is intractable for robots to have all functionalities pre-programmed prior to deployment. Instead, it is more realistic for robots to engage in supplemental, on-site learning about the user's needs and preferences and particularities of the environment. As a result, robots require the ability to adapt to and learn from their users. This learning may occur in the presence of or involve the user, and the observation of the robot learning may have an impact on the end-user's perceptions of the robot. In this work, we investigate the impacts on end-users of in situ robot learning through a series of human-subjects experiments. We investigate how different learning methods influence both in-person and remote participants' perceptions of the robot. While we find that the robot's learning method impacts perceived anthropomorphism ($p = .001$), we find that it is the participants' perceived success of the robot that impacts the participants' trust ($p < .001$) in and perceived usability of the robot ($p < .027$) rather than the robot's learning method. Therefore, when presenting robot learning, the performance of the learning method appears more important than the method of learning itself. Furthermore, we find that physical presence impacts perceived safety ($p < .001$), trust ($p < .001$), and usability ($p < .027$). Thus, for tabletop manipulation tasks, researchers should consider the impact of physical vs. virtual interactions for experiment participants.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *Accessibility theory, concepts and paradigms.*

KEYWORDS

robotics, robot learning, learning from demonstration, reinforcement learning, trust, human robot interaction, care robotics

ACM Reference Format:

Anonymous. 2018. Impacts of Robot Learning on User Attitude and Behavior. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The world's population is aging, and many older adults may require physical and cognitive assistance as they age. Yet, there is a growing shortage of nurses, caregivers, and facilities, inducing frequent caregiver burnout [39, 59]. One solution is to develop in-home care robots that can provide assistance and enable aging-in-place. However, deploying a robot that is fully capable into an unstructured home environment is difficult. Different users may want the same task done differently, and the way these tasks are done may need to change over time as the user ages or the environment changes. Thus, it can be beneficial for robots to perform behavior customization in situ. Users may additionally want to teach functionality to their robot so it can learn new tasks on-the-fly.

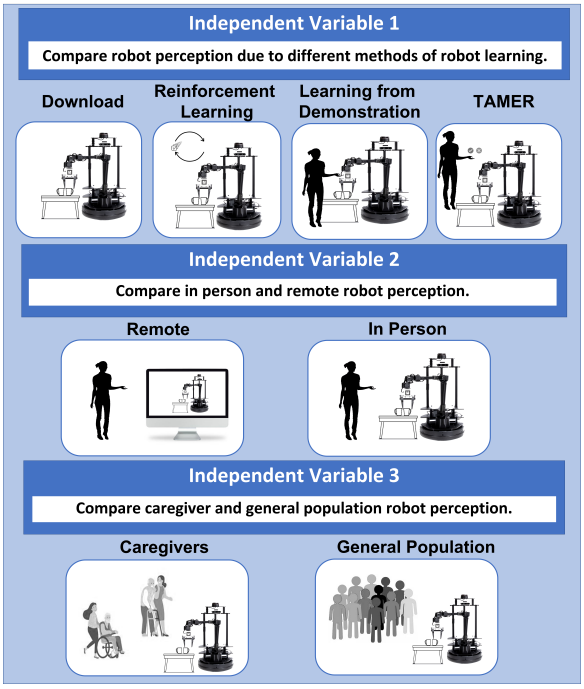

Figure 1: This figure shows the three factors we investigate.

Because the robot's learning will likely be visible to the end-user, and can even involve the user, the way in which the robot learns may have an impact on the end-user's perception of the robot. Different learning methods will require the robot to interact with the environment and the end-user in differing ways. For example, some learning methods will require the robot to explore the environment and fail many times before succeeding. Similarly, some learning methods will require the end-user to directly interact with the robot. These differences between robot learning methods may alter an end-user's perception of the robot and affect adoption.

In this work, we investigate how observing robot learning impacts an end-user's perception of a robot. We explore three factors which we hypothesize may impact perception of a robot during in situ learning. The first factor we investigate is learning condition. Our objective is to determine how the way in which a robot learns impacts an end-user's perception of the robot. We investigate a "download," control condition where the robot downloads a skill and the user does not see learning, two forms of learning from demonstration (LfD), and a reinforcement learning (RL) condition.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

We evaluate these learning methods as they are popular learning methods across a range of human interaction modes.

The second factor we investigate is differences in robot perception between the caregiver population and the general population. Caregivers represent a range of roles from nurses at an assisted living facility to family members and friends of a person needing assistance. Depending on the situation, caregivers may be the ones observing and guiding the robot's learning and deciding when the robot is safe and trustworthy for use. Additionally, in an in-home setting, a caregiver may also need to live with the robot. We conduct our study on both the general population and the caregiver population to determine the difference in robot perceptions between caregivers and the general population. Due to COVID-19 and the burden of in-person studies, we chose to conduct studies with caregivers virtually. Thus, to control for confounds, we employ a third factor: physical presence of the participant. For this factor, we investigate the differences for an end-user observing a robot learning in-person versus remotely.

In our work, we contribute the following:

(1) We conduct a human subjects experiment with three factors relevant to in situ learning: physical presence, population, and interaction mode for robot learning.
(2) We investigate how robot learning method impacts a user's perception of the robot and find that observing the robot learning impacts perceived anthropomorphism ($p = .002$).
(3) We examine how the general population's attitudes differ from that of caregivers, finding that caregivers are more critical of robot success on medicine-related tasks ($p = .033$).
(4) We find that in-person participants report more favorably than remote participants on perceived safety ($p < .001$), reliability ($p < .001$), ease of use ($p = .005$), attitude ($p < .001$), intent to use ($p = .014$), and trust in the robot ($p < .001$). This highlights the need for researchers to consider physical presence in study design.

2 RELATED WORKS

In this section, we describe prior literature on care robots, acceptance, reliance, and physical presence.

2.1 Care Robots

Care robots are defined by their function to support care-givers and/or care-receivers [62]. With the growing aging population and the predicted shortage in caregivers worldwide [58], the need for care robots is expected to increase. Indeed, recent surveys show that both the older adults and their caregivers are in favor of using care robots [38, 44, 17]. Our work seeks to provide guidance in developing *in situ* robot learning modes for widespread deployment of care robots across different users.

Care robots perform a variety of assistive tasks and promote extended independent living generally, with functionality falling under the category of physical or medical assistance [28, 54, 18, 37]. Physical assistance includes activities such as navigation, fall-prevention, object manipulation, and household chores [51, 20, 25, 31, 35, 42]. Medical assistance includes tasks such as health monitoring, medicine delivery, and the exertion of a social presence for coaching or social interaction [63, 12, 61, 52, 53, 29].

Most prior work in human-robot interaction (HRI) has investigated designing care robots to mitigate physical and cognitive decline in older adults [31, 23, 48] and factors that influence the adoption of care robots [6, 8]. However, these care robots are generally described as being deployed fully capable. Such an assumption is impractical for large-scale deployment across diverse users. Indeed, in prior work on user preferences for physically assistive robots, caregivers explicitly request robots that can learn to adapt to their user [9]. To the best of our knowledge, our work is the first to investigate how different learning techniques can influence the user's perceptions of a care robot.

2.2 Acceptance, Trust, and Reliance

Acceptance of care robots depends not only on the benefits the robot can bestow upon the user, but also on the user's perception of and attitudes towards the robot [10]. In this study, we employ the Technology Acceptance Model (TAM) survey to measure perceived usefulness, perceived ease of use, attitude, intention to use, and trust [7]. Additionally, we measure situational trust, resulting from an interaction with the robot, which is critical in understanding user acceptance [66, 36] and reliance on robots and is based upon situational uncertainty or vulnerability [33, 60]. Proper trust calibration mitigates over-compliance and under-reliance. However, prior work has yet to study how trust is impacted by the observation of learning, and how this observation may impact inappropriate reliance [14]. We investigate these research questions in our work.

Prior work has shown that a robot's performance has a significant impact on trust [24, 50, 57]. Salem et al. showed that trust and user reliance are negatively correlated with robot failure [55]. Hedlund et al. demonstrated that, when teaching a robot with LfD, if the robot failed people trusted the robot and themselves less [26]. Research also suggests that robot failure may make the robot more likable [41]. Further, the timing of the failure also impacts trust [13].

2.3 Anthropomorphism and Physical Presence

Prior work has shown that physical presence impacts a user's experience in robot interactions. Wainer et al. showed that the presence of a co-located physical robot improved user experience and perceived social awareness in a game [3]. Powers et al. found that users tend to engage longer with a co-located robot [45]. Gombolay et al. showed that users relied more appropriately on physical robots than on computer-based decision support systems in a scheduling task [22]. Contrarily, Natarajan et al. investigated how various robot behaviors and robot embodiments can influence trust in and reliance on robotic decision-support, finding no difference between virtual and physically present robots for a math problem task where the robot provided advice [43]. Therefore, the impact of physical presence may be task dependent. Our experiment provides evidence for how physical presence impacts a manipulation type task.

As the physical presence of a robot has been shown to impact the quality of interaction as perceived by the user, we include physical condition as a factor in our work [45, 22]. Moreover, prior work has also shown that the physical presence of robots can lead to increased perceived anthropomorphism [30]. Anthropomorphism can affect trust as well as other user attitudes such as user empathy

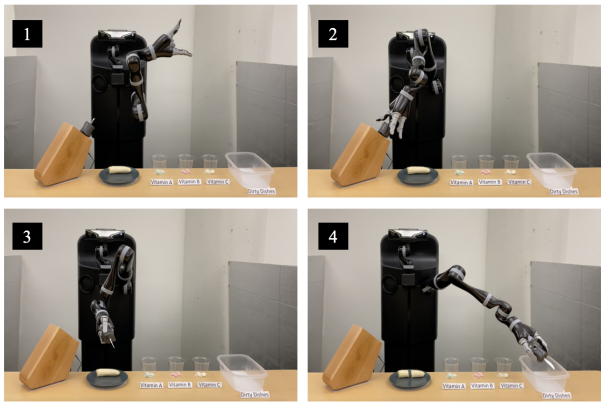

Figure 2: This figure shows our study set up and a sample training trajectory for the cutting sub-task.

[49, 43]. Hence, we measure perceived anthropomorphism in our study, evaluating it as a confound and as a dependent variable.

3 METHODOLOGY

This study consists of two phases. First, in the training phase, the participant observes videos of the robot learning a training task. Then, the participant observes the final performance achieved from the learning. Second, in the testing phase, the participant observes, and can intervene, as the robot uses what it has learned to accomplish a similar, riskier test task. In this section we describe the experimental domain, independent variables (IVs), research questions, metrics, and study procedure.

3.1 Domain

The domain we choose is a plate-making task, depicted in Figure 2. The robot must assemble a breakfast plate by preparing a breakfast item and placing it on the plate along with the proper medicine. We choose this domain in order to evaluate our work on both a manipulation task and a user-specific cognitive task. Plate-making involves the manipulation aspect of picking up a knife and cutting a banana in half. It additionally involves the cognitive preparation aspect of picking the right medicine and pouring it on the plate. This domain is motivated by prior work, showing a need for medicine dispensing and manipulation tasks [17, 46, 64]. The sub-tasks that compose this domain include picking up the knife, cutting the banana, placing the knife in the dirty dishes bin, picking up the proper medicine cup, pouring the medicine onto the plate, and placing the cup in the dirty dishes bin.

To evaluate trust, the task must create a sufficient level of uncertainty or imposition of vulnerability on the participant [33, 60]. As a result, we heighten the stakes in the test task to ensure that participants consider their risk tolerance when evaluating the robot. While training, the knife used is a small plastic knife, and the medicine dispensed is composed of different types of vitamins. During the testing phase, the knife becomes a large, sharp, metal knife, and the three medicines are labeled as morphine, aspirin, and antibiotics. This increased risk in the testing phases is inspired by the need for in-home robots to generalize to differing objects and environments. For example, one would conduct learning on a safe plastic knife when failure is expected during the learning process. However, success on novel variations of a task is not guaranteed with most learning systems, so we sought to include this uncertainty and the associated risk in our experiment.

3.2 Independent Variables

We describe the three independent variables of this study.

IV 1: Method of Robot Learning

We compare three visible learning conditions (RL, LfD, and Training an Agent Manually via Evaluative Reinforcement (TAMER)) with a baseline download condition where the robot learns but the learning is not visible. We aim to investigate how people may feel differently about the robot depending on the level of user-involvement in the robot learning. We note that in the LfD and learning from feedback conditions, the teacher is a good demonstrator, i.e. their demonstrations or feedback lead to the proper learned behavior.

Download: In the download condition, the participant observes the robot download the task knowledge from "the cloud." This serves as the control condition, as no learning is observed, but the robot does acquire, i.e. learn, the task. In this condition, the description of the robot is as follows: "Movo knows how to do various household tasks. You will watch as Movo performs them."

Reinforcement Learning (RL): Our no-involvement learning condition is the RL condition, where the robot learning happens independently from the user, based on some internal reward for the task. In this condition, the robot shows trial-and-error learning, iteratively learning sub-tasks until the overall task is completed. For the cut banana task for instance, first the robot learns to pick up the knife, then the robot learns to cut the banana. We choose to show this sequential sub-task learning in order to demonstrate that the agent is improving through incremental knowledge gains. No explicit reward function is explained, and we intentionally describe stages of the learning vaguely, rather than providing iteration count or time metric. In this condition, the description of the robot is as follows: "Movo does not know how to do household tasks yet. You will watch as Movo learns to do them."

Learning from Demonstration (LfD): Our high-involvement learning condition is LfD, where a demonstrator actively shows the robot how to do the task. For this study, we chose kinesthetic teaching, where the demonstrator physically moves the robot through the task. We chose kinesthetic teaching since prior work has shown that users find kinesthetic teaching is easier to demonstrate than other methods, such as teleoperation [2]. Additionally, we posit that watching kinesthetic teaching would be intuitive for participants to understand. In the LfD condition, the same trial-and-error learning is observed; however, we intersperse a video of a human teaching the robot the sub-tasks, prior to improvement in performance on these sub-tasks. For instance, the robot tries to pick up the knife and fails. Then, the participant observes a demonstrator show the robot how to pick up the knife, followed by the robot moving on its own to pick up the knife. Next the robot attempts to cut the banana. In this condition, the description of the robot is as follows: "Movo does not know how to do household tasks yet. You will watch as Movo's owner demonstrates how to do these tasks."

TAMER The last condition is the TAMER [32] condition, which is a middle ground between LfD and RL. In this condition, the robot attempts the task on its own and the user provides binary feedback throughout the learning process to shape the robot's behavior. In this condition, the same trial-and-error learning is observed; however, we add a visual of a human teaching the robot through positive or negative (i.e., binary) feedback. This feedback is shown to the user through a remote control with green and red buttons pressed throughout the training to convey positive and negative feedback, respectively [7]. In this condition, the description of the robot is as follows: "Movo does not know how to do household tasks yet. You will watch as Movo's owner teaches Movo to do them."

IV 2: General Population vs Caregiver
We compare robot perceptions between the general population and the caregiver population.

General Population: We invite members adjacent to a metropolitan university campus to take our study.

Caregiver Study: We invite our target population of caregivers in our state to take our study. We define caregivers as persons who care for members of the community that require assistance in their day-to-day life. This definition encompasses a variety of care providers, ranging from family and friends that care for a loved one to employees of assisted living residences such as gerontological nurses and physical therapists.

IV 3: In-Person vs Remote
We compare robot perception between in-person and remote participants of our study.

Remote Study: We invite participants to take the study remotely, where the training and testing phase of the study are both conducted via a web application.

In-Person Study: We invite participants to take our study in person, in the laboratory. These participants engage in the training phase via the web application, and witness the test phase in person.

3.3 Research Questions

RQ 1 - *How does observing robot learning impact a user's perception of the robot? Furthermore, how does observing different methods of robot learning impact a user's perception of the robot?* We first want to understand how witnessing robot learning may impact a user's perception of the robot, and subsequent adoption of and reliance on the robot. We additionally investigate how different methods of learning may impact a user's perception of the robot.

RQ 2 - *How does the caregiver population's perception of learning robots differ from the general population?* Prior work that has shown that caregivers feel favorably regarding robot usability and intent of future use [38, 44]. We seek to understand how people in general feel about robots learning and how this may differ from how care givers feel about robot learning.

RQ 3 - *How does a robot's physical presence impact a user's perception of the robot?* COVID-19 limits our in-person interactions with care givers, so we conduct the study with care givers remotely. Prior work has shown physical presence to impact user perception [45]. In order to avoid confounds of caregivers being remote and general population being in person, we evaluate the general population both remotely and in person.

3.4 Metrics

First, we describe the metrics collected in our Pre-Study, Post Trial, and Post Study Questionnaires. In addition to the surveys below, we ask the participants two, ad hoc questions in the Post Study Questionnaire. First we ask the participant what tasks – from a list of hand-crafted tasks both in and outside of the distribution of tasks observed in the study – they would trust the robot to do. This question measures the extent of user adoption. We also ask an open-ended question about the participant's understanding of the robot's learning to qualify their perception of robot competence. All surveys comply with the guidelines detailed in Schrum et al., when possible [56].

3.4.1 Pre-Study Metrics.

Demographic Information: We collect participant education [47], parental education, and participant age and gender. We additionally collect the field of occupation of the participant, using the ACT's List of College Majors and Occupational Choices [1].

Personality: We employ the Mini-IPIP, a 20-item, 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), to measure the following five personality traits: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness [15].

Computer and Robotics Prior Experience We evaluate the participant's prior experience with computers using the Computer Usage Checklist 6-item, 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) found in Correlates of Computer Anxiety in College Students [47]. We additionally alter this scale to create a new, hand-crafted robotics experience scale (see Appendix).

Dispositional Trust We evaluate baseline participant trust using the 6-item, 5-point Propensity to Trust Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) developed by [40].

3.4.2 Post-Trial Metrics.

Perceived Success: Participants report whether they believe the robot was successful after each trial (binary metric).

Perceived Safety: We employ Godspeed's perceived safety scale 3-item, 5-point semantic differential scale [5].

Reliability: We employ the trust in automation subscale for reliability and competence, a 6-item, 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) [34].

Intervention (Reliance): Human intervention is a commonly used technique in HRI to assess users' reliance in an autonomous agent such as a robot [11, 4]. In our study, we ask participants to interrupt the robot during a trial if they feel unsafe or uncomfortable, if the robot makes a mistake, or if they anticipate the robot might make a mistake. We use interruption as an objective metric to measure a participant's reliance on the robot.

3.4.3 Post-Study Metrics.

Anthropomorphism: We subjectively measure the participants' perception of the robot's anthropomorphism through Godspeed's anthropomorphism scale, a 5-item, 9-point semantic differential scale [5]. Note that we keep the robot consistent between conditions.

Usability and Acceptability: We employ the technology acceptance model (TAM) 15-item, 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) from Integrating trust and personal values into the Technology Acceptance Model [7]. We disregard the two questions regarding environmental and time concerns, as these did not apply to our domain. We report our findings with regards to the five TAM sub-scales: usefulness, ease of use, attitude, trust, and intent to use. We do not analyze results related to the trust subscales as this information is already captured in the Situational Trust scale described below.

Situational Trust: To determine trust in the robot after observing performance, we collect perceived situational trust using the Checklist for Trust Between People and Automation, a 12-item, 7-point Likert-like scale (1 = Not at all to 7 = Completely) [27].

Task Risk: We employ a 9-item, 9-point semantic differential scale to qualify the risk of a task [19].

3.5 Human-subjects Experiment

3.5.1 Procedure. Our experiment was approved by an Institutional Review Board (IRB). Participants start by watching a video in which the robotic agent introduces itself and demonstrates its range of mobility and degrees of freedom. After watching the unboxing video, participants fill out the pre-study questionnaire. Then, participants go through the training phase where they observe the robot learn to perform the task. Participants are told to "imagine that this learning occurs in your home." The learning lasts a total of approximately 10 minutes, with interspersed attention checks.

Ours is a between-subjects experiment where participants are broken into four learning conditions (i.e., download, RL, LfD, and TAMER) and two physical conditions (in-person, remote) (eight total combinations).The difference between learning conditions is the type of robot learning observed. The difference between physical conditions is whether the testing trials were videos or conducted in-person on the MOVO Beta robot.

In the training phase, participants watch their condition's unique training video, where the cutting task is performed using a plastic knife and the medications are different types of vitamins. After this training phase, all participants observe the same final training performance video; therefore, robot end performance is held constant.

Next, participants go through the testing phase. Participants observe the testing trials where the robot states its goal and demonstrates a success rate of about 80%. The testing tasks are performed using a metal knife and the medications are labeled as antibiotics, aspirin, and morphine. The risk during the testing task is intended to be higher to evaluate how participants trust the robot during high-risk situations. During each testing trial, the participants are instructed to interrupt the robot if they feel unsafe or uncomfortable, if the robot makes a mistake, or if they anticipate the robot might make a mistake. The remote and caregiver participants view the testing tasks as videos in the web application with an intervene button on the screen, whereas the in-person participants watch the physically present robot and press the red button on a gaming controller to interrupt. The interruption data collected here serves to assess reliance.

Participants first have one practice trial for familiarization with the intervention mechanism to interrupt the robot. Then, participants view nine testing trials: two successful attempts by the robot in the both the cutting and medicine preparation domain, one ambiguous success in both of these domains, one failure in both of these domains, and one success trial that incorporates both the cutting and medicine. We choose the success rate for robot task completion to be $\approx$ 70% based upon previous user studies examining trust and reliance [65, 67]. This success rate shows that the robot is capable of completing the task, as was observed at the end of training. Yet, we also want the robot to fail a few times to be able to test for participant over-reliance. Nine trials balances the need to exhibit a variety of task trajectories and obtain repeated measures while not fatiguing the participant.

After each of the nine testing trials, participants fill out the Post-Trial Questionnaire where we ask them to rate the degree to which they trust the robot to act safely and the degree to which they believe the robot will accomplish the task. The testing iterations are shown in person, in real time, on the physical robot for the in-person participants and through recorded videos of identical robot behavior for the remote participants. After the testing phase, the participants complete the Post-Study Questionnaire. We have provided videos of the training and testing phase in the Supplementary.

3.6 Design Decisions

Policies: A robot will fail during the learning process. Because prior work has shown that failure impacts trust and impression of the robot [55], in order to prevent failure from being a confound in our study, we choose to keep performance and order of trials consistent between conditions to isolate the impact of observing the learning process. To do so, we demonstrate our learning conditions via Wizard-of-Oz policies rather than training RL, LfD, and TAMER models. This enables us to keep agent behavior, training time, and failure rate consistent and allow us to isolate the observation of learning as the IV.

Ambiguous Trials: The testing phase includes nine trials with five successes, two failures, and two ambiguous trajectories. The ambiguous trajectories are ultimately successful in accomplishing the task goal, but are intended to be unclear whether the robot will succeed at certain points. Inspired by robot legibility by Dragan et al. [16], in the ambiguous trial with the medicine task, the robot first moves as if it might pick up the wrong medication and then moves as if it is deciding between the incorrect and correct medication. In the cutting task, the robot first angles the knife point down towards the table and takes a very roundabout route to reach the banana. This behavior was designed to test whether participants would intervene during ambiguous trials.

4 RESULTS

We report results from 120 (60 in-person, 60 remote) participants from a college campus and 11 caregiver participants from assisted living facilities within a 50-mile radius of the campus for this study (55.73% female, 44.27% male, 65.65% aged 18-24, 29% aged 25-34, 3.05% aged 45-54, 2.29% aged 55 and older) . Each participant was

compensated $25 for completing our study.[1] We randomize each participant's learning condition, ensuring that we counterbalance the quantity of participants per condition.[2] For the general population, we further randomize each participant's physical presence condition.

In our analysis, we evaluate each dependent variable described in Section 3.4 with respect to our IVs and covariates. Here the IVs are learning condition, physical condition or population. Before applying a parametric test, we ensure that the data is normally distributed using the Shapiro-Wilk normality test and we test for homoscedasticity using Levene's Test for Homogeneity of Variance. If the data does not pass normality and homoscedasticity or the data is ordinal, such as count data, we use a non-parametric test. All statistical tests, model details, and tests for assumptions are reported in the Appendix. Significance is measured as $\alpha < .05$. We report Cohen's d effect size as d.

We note that there are minor differences between the way the remote and in-person populations experienced the robot's performance due to technical difficulties for the in-person experiment with a live robot performing the tasks. For example, pills occasionally spilled off the plate in a trial that was meant to be successful. However, we find that there is no statistically significant difference between the perceived success comparing in-person versus remote study conditions. While this result is positive for our study design, we further mitigate any potential confound by including perceived success as a covariate in all other models.

4.1 Analysis of RQ 1: Learning Methods

We first investigate the differences in robot perception between our various learning conditions (i.e., Download, RL, LfD, and TAMER). We conduct this analysis on the combined data from both the physical and virtual participants. We present the following findings.

Anthropomorphism: A Kruskal-Wallis (KW) test for anthropomorphism shows a main effect for learning condition ($\chi^2(3) = 16.241, p = .001$). We then determine which learning condition pairs are significantly different. As shown in Figure 3, a Wilcoxon rank sum test with Bonferroni correction shows a statistically significant difference in perceived anthropomorphism between the following learning conditions: Download vs RL ($W = 325.5, p = .019, d = -.663$), Download vs TAMER ($W = 323, p = .044, d = -.515$), and RL vs LfD ($W = 775.5, p = .018, d = .681$).

When we evaluate the anthropomorphism in in-person and remote participants separately, we find no significant difference in anthropomorphism between learning conditions for in-person participants. For remote participants, we find significance ($\chi^2(3) = 18.731, p < .001$). The relevant pairs of learning conditions are as follows: Download vs RL ($W = 18.5, p < .006, d = -1.699$), Download vs TAMER ($W = 37, p = .011, d = -1.330$), and RL vs LfD is trending significant ($W = 172, p = .084, d = .919$). We find that RL is higher than Download and LfD, and that TAMER is higher than Download.

[1]Due to some facility guidelines, some caregivers were not monetarily compensated. This lack of compensation was approved by the IRB as per facility requirements.
[2]The caregiver population was not adequately counterbalanced between learning conditions; however, we do not report on learning condition for this population.

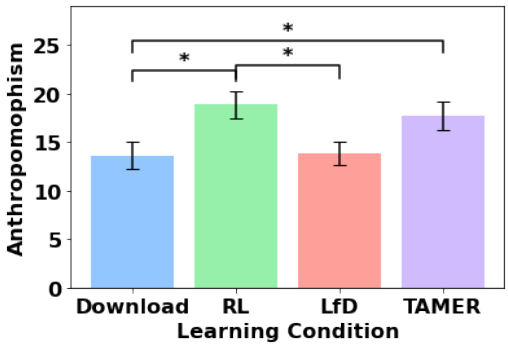

Figure 3: Anthropomorphism for the learning conditions.

We did not observe statistically significant differences in trust in the robot, perceived risk of the task, and perceived usability of the robot between learning conditions. Thus, our results suggest that level of human involvement of the learning condition in our study did not strongly impact user trust the robot, how usable the robot appears to them, or how risky the task is.

4.2 Analysis of RQ 2: Caregiver

We first investigate whether there is a difference in perceived success between the caregiver and general populations. We employ a KW test to determine whether there is a difference between the caregiver population and the general population for perceived success on the medicine and cutting sub-tasks. We find significance for the perceived success on the medicine sub-task ($\chi^2(1) = 4.5454$, $p = .033, d = .530$), indicating that caregivers rate the robot's success on medicine sub-task as lower than the general population. However, we do not find that caregivers interrupt more on the medicine trials than the general population. We hypothesize that caregivers are more wary of robot success on the medicine sub-task given their experience compared to our general population pool.

Next, we run the KW test on both computer and robotics experience, and find a main effect with population for both ($\chi^2(1) = 14.316, p < .001, d = 1.491$) and $\chi^2(1) = 15.583, p < .001, d = 1.211$) respectively, with the caregiver population reporting lower prior experience in both computer and robotics experience. Additionally, we note that a KW test shows agreeableness is trending towards significance with respect to population ($\chi^2(1) = 2.832, p = .092, d = .493$), with caregiver population demonstrating lower agreeableness than the general population. A KW test shows that propensity to trust is trending towards significance with population ($\chi^2(1) = 3.0399, p = .081, d = .526$) with caregivers reporting lower propensity to trust than the general population.

Finally, we run an ANOVA on the linear regression model for the risk scale, which was trending towards significant ($F(1, 129) = 3.583, p = .061, d = -.596$). While not significant, we note this result that it appears as though caregiver participants find the task more risky than their general population counterpart. We plan to recruit additional participants to increase the test's power.

4.3 Analysis of RQ 3: Physical Presence

Perceived Safety: We run a KW test on perceived safety and find a main effect for physical condition ($\chi^2(1) = 11.55, p < .001, d =$

$-.621$). Notably, we find that in-person participants report higher perceived safety compared to remote participants. This suggests that studies conducted remotely can expect participant's perceived safety to increase if conducted in person. We posit this could be due to the feeling of safety engendered from taking part in an IRB-approved user study. Alternatively, this could be due to the physical presence of the researcher, which prior work has shown impacts a participant's social behavior [21].

Reliability: We ran an ANOVA for reliability [34] and found that participants report higher reliability for the in-person condition (($F(1, 129) = 19.2537, p < .001, d = -.602$) compared to the remote condition. Further, we find that interruption count (($F(1, 129) = 6.6873, p = .011, d = 8.192$), perceived success (($F(1, 129) = 41.9687, p < .001, d = 6.565$), conscientiousness ($F(1, 129) = 3.9242, p = .05, d = 2.513$), and anthropomorphism (($F(1, 129) = 23.1017, p < .001, d = .999$) positively impact reliability.

Situational Trust: Next, a linear regression model of trust [27] showed a main effect for physical condition ($F(1, 129) = 22.8530, p < .001, d = -.670$), with participants in the remote condition reporting lower trust compared to the in-person participants. We find that perceived success (($F(1, 129) = 26.3919, p < .001, d = 3.845$) and anthropomorphism (($F(1, 129) = 30.8040, p < .001, d = 2.335$)) positively impact trust.

Usability and Acceptance: We ran an ANOVA on the linear model for each sub-scale of the TAM model [7]. An ANOVA showed that in-person participants rated the ease of use sub-scale more favorably than remote participants ($F(1, 129) = 8.1177, p = .005, d = -.441$). We found that ease of use was positively impacted by perceived success ($F(1, 129) = 11.7037, p < .001, d = .997$) as well as anthropomorphism ($F(1, 129) = 24.8910, p < .001, d = -1.262$). An ANOVA on the attitude sub-scale showed that in-person participants had a more positive attitude than remote participants ($F(1, 129) = 21.7933, p < .001, d = .652$). Perceived success ($F(1, 129) = 12.4402, p < .001, d = 1.926$), openness ($F(1, 129) = 5.0349, p = .027, d = -.981$), propensity to trust ($F(1, 129) = 10.1553, p = .002, d = -2.849$), and anthropomorphism ($F(1, 129) = 17.0098, p < .001, d = -.664$) positively impacted attitude. For the intent-to-use sub-scale, we find significance with a KW test ($\chi^2(1) = 6.0633, p = .014, d = -.434$), with in-person participants reporting higher intent to use than remote participants.

In summary, we find significance with the ease of use, attitude, trust, and intention to use sub-scales (no significance for the usefulness sub-scale) between physical conditions, with in-person participants reporting more favorably on all three scales. This might be the case because in-person participants interact with the robot and can therefore imagine using it more concretely. Additionally, we found that anthropomorphism and perceived success significantly impacted the ease of use and trust sub-scales of the TAM.

Risk: We ran an ANOVA on the linear model for risk and found the physical condition to be significant ($F(1, 129) = 4.4668, p = .037, d = .352$), with remote participants reporting higher risk than their in-person counterparts. Additionally, we find that conscientiousness ($F(1, 129) = 7.3099, p = .008, d = 2.666$) positively impacts risk and propensity to trust ($F(1, 129) = 8.5143, p = .004, d = 1.554$) negatively impacts risk.

Interruption Count: Finally, we compare the number of times participants interrupted between the physical conditions. With the

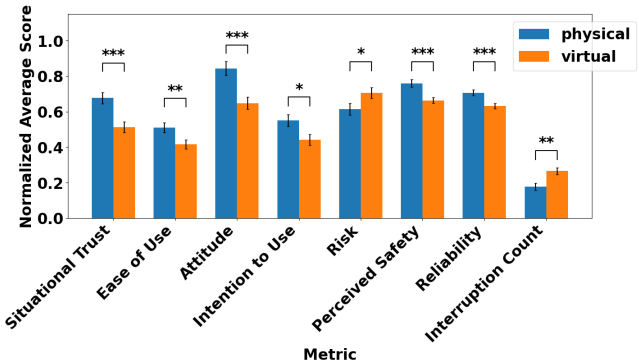

Figure 4: Significant metrics for in-person vs. remote.

KW test, we find that remote participants interrupt more than their in-person counterparts ($\chi^2(1) = 10.315, p = .001, d = .550$). We hypothesize this is because of the lower trust and perceived safety remote participants feel with regards to the robot.

To examine when people decided to interrupt the robot, we compared how much they interrupted to their perceived success of the robot over the course of all the trials. Through a Spearman's correlation test, we find a negative correlation between interruption count and perceived success ($p < .001, \rho = -.705$). However, we would have anticipated this correlation being stronger because we expected people to interrupt the robot when it failed to accomplish the task. Anecdotally, we found that in-person participants noticed the robot fail, often verbally expressing surprise at the failure, but did not interrupt the robot. Instead, they let the trajectory play out, curious to see what would happen. We question whether these participants might have intervened more if this study took place in their home instead of a laboratory setting.

Next, we found a main effect with the Friedman rank sum test comparing interruption count and ground truth success with subject ID as the repeated measure ($\chi^2(2) = 144.54, p < .001$). We find the relevant significant pairs with a Nemenyi post-hoc to be that participants significantly interrupted failure more than success ($p < .001, d = 1.943$) and failure more than ambiguous ($p < .001, d = -1.349$), both of which make intuitive sense. We also observe that interruption count in ambiguous trials is trending towards being significantly more frequent than interruption in successful trials ($p = .054, d = .541$).

A Spearman's rank correlation ($p = .020, \rho = .203$) showed that remote participants who are more agreeable are more likely to interrupt successful trials. This finding could be due to experimenter expectancy effect, i.e. the participant's belief that we wish for them to interrupt the trial. Furthermore, a Spearman's rank correlation ($p = .034, \rho = -.254$) showed that the frequency with which remote participants interrupt ambiguous trials is negatively correlated with their prior robot experience. This result suggests that people that are more familiar with robots hold off intervening when the robot's performance is ambiguous.

Additionally, a Friedman rank sum test between interrupt frequency and sub-task type was significant ($\chi^2(1) = 21.053, p < .001, d = .523$), showing that trials that involve the cutting sub-task were interrupted more than the medicine sub-task. We suspect that this is due to the difference in relative permanence of the mistakes.

One participant noted they did not interrupt the medicine sub-tasks as the error was something they could fix, but that they interrupted the cutting sub-tasks as these were unrecoverable failures.

5 DISCUSSION

RQ 1: Learning Methods – RQ1 investigates the differences in perceptions of the robot due to different methods of learning demonstrated to the user. We find that participants tend to perceive the RL condition as more anthropomorphic than the Download ($p = .019$) and LfD conditions ($p = .018$) and that TAMER is perceived as more anthropomorphic than the Download condition ($p = .044$). This suggests that if the participant is observing a learning condition, the more human intervention and guidance is observed, the less anthropomorphic the agent is perceived to be. This may be because, in the absence of a clear, human-like method of learning (via demonstration of with visible feedback), the participant assigns a "personality" to the robot. Since anthropomorphism also impacts trust and ease of use, one should consider how anthropomorphic the robot may appear when selecting a learning algorithm.

Overall, we find that perceived success impacts reliability ($p < .001$), situational trust ($p < .001$), and ease of use ($p < .001$), but learning condition does not. We thus posit that users may prioritize choosing a robot learning method that is more competent over a latent user preference over interaction modalities.

RQ 2: Caregiver Population – RQ2 investigates the differences between caregiver and general population perceptions of the robot. We find that caregivers rate the robot's success on the medicine sub-task as lower than the general population ($p = .033$). We hypothesize that given their background, caregivers are more skeptical of robot success on the medicine sub-task than the general population. However, this does not appear to translate to lower reliance.

We find that the caregiver population has lower prior experience with computers ($p < .001$) and robots ($p < .001$). This could result in caregiver participants not feeling qualified to interrupt the robot. We also note trending lower agreeableness in our caregiver population ($p = .092$). This is notable, as we found agreeableness to be positively correlated with interruption rate for remote participants ($p = .020$). Thus, we posit that with a larger caregiver population sample, results might trend towards lower interruption counts. Finally, we see preliminary evidence that perceived risk ($p = .061$) and propensity to trust ($p = .081$) are trending towards significance with the caregiver population reporting higher risk and lower propensity to trust than the general population.

RQ 3: Physical Presence – RQ3 examines the impact of physical presence on a user's perception of morebot. We find that in-person participants report more favorably with regards to perceived safety ($p < .001$), reliability ($p < .001$), ease of use ($p = .005$), attitude ($p < .001$), intent to use ($p = .014$), and trust in the robot ($p < .001$) than remote participants. Additionally, in-person participants report lower risk ($p = .037$) and interrupt the robot less frequently ($p = .001$) than remote participants.

We hypothesize that in-person participants may feel more safe, and more trusting in the robot and the experiment in general, due to the presence of the researcher throughout the study, and due to the IRB-approved status of the user study. Furthermore, the in-person participants observe the robot's performance in real time, in close physical proximity to the robot, potentially increasing the participants' ability to imagine using the robot, and therefore the perceived usability of the robot. We additionally note that when the in-person participant intervenes in the robot's behavior, they observe it stop, and observe the researcher reset the robot to it's home position whereas in the remote condition the video simply ends and is removed from the screen. This difference in outcome may play a role in the perceived usability, trust, and safety, as the outcome of an intervention is better understood by in-person participants. The lower levels of interruption with in-person participants should be taken into consideration if researchers rely on participants to intervene in cases of dangerous or faulty robotic behavior.

6 LIMITATIONS AND FUTURE WORK

A limitation of our work is the sample size ($n = 11$) of our caregiver population. Additionally, our participants from the general population were mostly college-aged students from a university campus. In the future, we would like to conduct this study on a larger population of caregivers and with a more diverse general population. Another limitation is the presence of the researcher standing near the emergency stop button, for in-person participant pool. This safety precaution is not necessary for the remote participants where the participant observes videos of the robot. This difference may introduce a potential confound. An additional limitation for the in-person participant pool is the occurrence of occasional accidental errors, such as pills bouncing off the plate on an intended successful trial. In our analysis, we accounted for participants' perceived robot success and found no confounding difference of perceived success between conditions.

While we employed Wizard-of-Oz policies, we propose to investigate people's trust in different methods of learning when learning actually occurs. A second avenue of future work involves trust calibration. As learning agents may not learn quickly and may fail often, it is important to evaluate the best method of trust calibration for embodied, learning agents [43]. Thus, we propose to determine the techniques that are most effective in repairing trust in trust-deficient learning conditions. Finally, we aim to investigate differences between user perceptions of a robot in an in-person laboratory setting compared to a study conducted in the home.

7 CONCLUSION

In this work, we conduct a series of human-subjects experiments to assess user attitudes towards the concept of embodied care robots that learn in the home, as compared to robots that are delivered fully-capable. We investigate the impact of physical presence and learning condition on robot perception. We additionally compare the impact of these factors on the caregiver population, as compared to the general population. We find that the robot's learning method impacts perceived anthropomorphism ($p = .001$) and that caregivers are more wary of the robot performing medicine dispensing tasks compared to the general population. Finally, for table-top manipulation tasks, physical presence favorably impacts perceived safety ($p < .001$), trust ($p < .001$), and usability ($p < .027$).

REFERENCES

[1] ACT. 2015. College Majors and Occupational Choices. en. (2015). Retrieved Feb. 14, 2022 from https://www.act.org/content/act/en/research/reports/act-publications/college-choice-report-class-of-2013/college-majors-and-occupational-choices/college-majors-and-occupational-choices.html.

[2] Baris Akgun and Kaushik Subramaman. 2011. Robot learning from demonstration: kinesthetic teaching vs. (2011).

[3] Wilma A. Bainbridge, Justin Hart, Elizabeth S. Kim, and Brian Scassellati. 2008. The effect of presence on human-robot interaction. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 701–706. DOI: 10.1109/ROMAN.2008.4600749.

[4] Anthony L. Baker, Elizabeth K. Phillips, Daniel Ullman, and Joseph R. Keebler. 2018. Toward an understanding of trust repair in human-robot interaction: current research and future directions. *ACM Trans. Interact. Intell. Syst.*, 8, 4, Article 30, (Nov. 2018), 30 pages. DOI: 10.1145/3181671.

[5] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. en. *International Journal of Social Robotics*, 1, 1, (Jan. 2009), 71–81. DOI: 10.1007/s12369-008-0001-3.

[6] Jenay M Beer, Cory-Ann Smarr, Tiffany L Chen, Akanksha Prakash, Tracy L Mitzner, Charles C Kemp, and Wendy A Rogers. 2012. The domesticated robot: design guidelines for assisting older adults to age in place. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 335–342.

[7] Daniel Belanche, Luis V. Casaló, and Carlos Flavián. 2012. Integrating trust and personal values into the Technology Acceptance Model: The case of e-government services adoption. en. *Cuadernos de Economía y Dirección de la Empresa*, 15, 4, (Oct. 2012), 192–204. DOI: 10.1016/j.cede.2012.04.004.

[8] Praminda Caleb-Solly, Sanja Dogramadzi, David Ellender, Tina Fear, and Herjan van den Heuvel. 2014. A mixed-method approach to evoke creative and holistic thinking about robots in a home environment. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 374–381.

[9] Gerard Canal, Guillem Alenyà, and Carme Torras. 2017. A taxonomy of preferences for physically assistive robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. (Aug. 2017), 292–297. DOI: 10.1109/ROMAN.2017.8172316.

[10] Amedeo Cesta, Gabriella Cortellessa, Maria Vittoria Giuliani, Federico Pecora, Massimiliano Scopelliti, and Lorenza Tiberio. 2007. Psychological Implications of Domestic Assistive Technology for the Elderly. *PsychNology Journal*, 5, (Jan. 2007), 229–252.

[11] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 307–315.

[12] S. Coradeschi et al. 2013. GiraffPlus: Combining social interaction and long term monitoring for promoting independent living. In *2013 6th International Conference on Human System Interactions (HSI)*. ISSN: 2158-2254. (June 2013), 578–585. DOI: 10.1109/HSI.2013.6577883.

[13] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 2013. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 251–258.

[14] Stephen R Dixon and Christopher D Wickens. 2006. Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Human factors*, 48, 3, 474–486.

[15] M Donnellan, Frederick Oswald, Brendan Baird, and Richard Lucas. 2006. The Mini-IPIP Scales: Tiny-yet-Effective Measures of the Big Five Factors of Personality. *Psychological assessment*, 18, (July 2006), 192–203. DOI: 10.1037/1040-3590.18.2.192.

[16] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. en. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Tokyo, Japan, (Mar. 2013), 301–308. ISBN: 978-1-4673-3101-2 978-1-4673-3099-2 978-1-4673-3100-5. DOI: 10.1109/HRI.2013.6483603.

[17] Laura Fiorini et al. 2021. Assistive robots to improve the independent living of older persons: results from a needs study. *Disability and Rehabilitation: Assistive Technology*, 16, 1, 92–102.

[18] Laura Fiorini et al. 2021. Assistive robots to improve the independent living of older persons: results from a needs study. *Disability and Rehabilitation: Assistive Technology*, 16, 1, (Jan. 2021), 92–102. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17483107.2019.1642392. DOI: 10.1080/17483107.2019.1642392.

[19] Baruch Fischhoff, Paul Slovic, Sarah Lichtenstein, Stephen Read, and Barbara Combs. 1978. How Safe Is Safe Enough? A Psychometric Study of Attitudes Toward Technological Risks and Benefits. *Policy Sciences*, 9, (Jan. 1978), 127–152. DOI: 10.1007/BF00143739.

[20] David Fischinger et al. 2016. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. en. *Robotics and Autonomous Systems*. Assistance and Service Robotics in a Human Environment 75, (Jan. 2016), 60–78. DOI: 10.1016/j.robot.2014.09.029.

[21] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. 2015. Systematic analysis of video data from different human–robot interaction studies: A categorization of social signals during error situations. *Frontiers in Psychology*, 6. Place: Switzerland Publisher: Frontiers Media S.A. DOI: 10.3389/fpsyg.2015.00931.

[22] Matthew Gombolay et al. 2018. Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research*, 37, 10, 1300–1316.

[23] Consuelo Granata, M Pino, G Legouverneur, J-S Vidal, P Bidaud, and A-S Rigaud. 2013. Robot services for elderly with cognitive impairment: testing usability of graphical user interfaces. *Technology and Health Care*, 21, 3, 217–231.

[24] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53, 5, 517–527.

[25] Swisslog Healthcare. 2013. Autonomous Service Robot For Hospitals - Savioke Relay. en-US. (2013). Retrieved Feb. 21, 2022 from https://www.swisslog-healthcare.com/en-us/products/transport-automation/relay-autonomous-service-robot.

[26] Erin Hedlund, Michael Johnson, and Matthew Gombolay. 2021. The Effects of a Robot's Performance on Human Teachers for Learning from Demonstration Tasks. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '21). Association for Computing Machinery, New York, NY, USA, (Mar. 2021), 207–215. ISBN: 978-1-4503-8289-2. DOI: 10.1145/3434073.3444664.

[27] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4, (Mar. 2000), 53–71. DOI: 10.1207/S15327566IJCE0401_04.

[28] David O. Johnson et al. 2014. Socially Assistive Robots: A Comprehensive Approach to Extending Independent Living. en. *International Journal of Social Robotics*, 6, 2, (Apr. 2014), 195–211. DOI: 10.1007/s12369-013-0217-8.

[29] Mari Kangasniemi, Suyen Karki, Noriyo Colley, and Ari Voutilainen. 2019. The use of robots and other automated devices in nurses' work: an integrative review. *International journal of nursing practice*, 25, 4, e12739.

[30] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition*, 26, 2, 169–181.

[31] Ralf Kittmann, Tim Fröhlich, Johannes Schäfer, Ulrich Reiser, Florian Weißhardt, and Andreas Haug. 2015. Let me Introduce Myself: I am Care-O-bot 4, a Gentleman Robot. en. In *Let me Introduce Myself: I am Care-O-bot 4, a Gentleman Robot*. De Gruyter, (Sept. 2015), 223–232. ISBN: 978-3-11-044392-9. DOI: 10.1515/9783110443929-024.

[32] W Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: the tamer framework. In *Proceedings of the fifth international conference on Knowledge capture*, 9–16.

[33] Spencer C. Kohn, Ewart J. de Visser, Eva Wiese, Yi-Ching Lee, and Tyler H. Shaw. 2021. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology*, 12. Retrieved Feb. 14, 2022 from https://www.frontiersin.org/article/10.3389/fpsyg.2021.604977.

[34] Moritz Körber. 2018. Theoretical considerations and development of a questionnaire to measure trust in automation. In (Mar. 2018).

[35] Ioannis Kostavelis, Dimitrios Giakoumis, Sotiris Malasiotis, and Dimitrios Tzovaras. 2016. RAMCIP: Towards a Robotic Assistant to Support Elderly with Mild Cognitive Impairments at Home. en. In *Pervasive Computing Paradigms for Mental Health* (Communications in Computer and Information Science). Silvia Serino, Aleksandar Matic, Dimitris Giakoumis, Guillaume Lopez, and Pietro Cipresso, (Eds.) Springer International Publishing, Cham, 186–195. ISBN: 978-3-319-32270-4. DOI: 10.1007/978-3-319-32270-4_19.

[36] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. 2019. Trust in socially assistive robots: Considerations for use in rehabilitation. en. *Neuroscience & Biobehavioral Reviews*, 104, (Sept. 2019), 231–239. DOI: 10.1016/j.neubiorev.2019.07.014.

[37] Hee Rin Lee and Laurel D. Riek. 2018. Reframing Assistive Robots to Promote Successful Aging. *ACM Transactions on Human-Robot Interaction*, 7, 1, (May 2018), 11:1–11:23. DOI: 10.1145/3203303.

[38] Jai-Yon Lee, Young Ae Song, Ji Young Jung, Hyun Jeong Kim, Bo Ram Kim, Hyun-Kyung Do, and Jae-Young Lim. 2018. Nurses' needs for care robots in integrated nursing care services. *Journal of Advanced Nursing*, 74, 9, 2094–2105.

[39] Meredith Mealer, Ellen L. Burnham, Colleen J. Goode, Barbara Rothbaum, and Marc Moss. 2009. The prevalence and impact of post traumatic stress disorder and burnout syndrome in nurses. eng. *Depression and Anxiety*, 26, 12, 1118–1126. DOI: 10.1002/da.20631.

[40] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward

Automation on Trust in an Automated System. en. *Human Factors*, 55, 3, (June 2013), 520–534. Publisher: SAGE Publications Inc. DOI: 10.1177/0018720812465081.

[41] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI*, 4. Retrieved Oct. 2, 2022 from https://www.frontiersin.org/articles/10.3389/frobt.2017.00021.

[42] Justinas Mišeikis et al. 2020. Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications. *IEEE Robotics and Automation Letters*, 5, 4, (Oct. 2020), 5339–5346. Conference Name: IEEE Robotics and Automation Letters. DOI: 10.1109/LRA.2020.3007462.

[43] Manisha Natarajan and Matthew Gombolay. 2020. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, 33–42. ISBN: 9781450367462. DOI: 10.1145/3319502.3374839.

[44] Maribel Pino, Mélodie Boulay, François Jouen, and Anne Rigaud. 2015. "Are we ready for robots that care for us?" Attitudes and opinions of older adults toward socially assistive robots. *Frontiers in Aging Neuroscience*, 7. Retrieved Feb. 14, 2022 from https://www.frontiersin.org/article/10.3389/fnagi.2015.00141.

[45] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 145–152.

[46] Pekka Rantanen, Timo Parkkari, Saija Leikola, Marja Airaksinen, and Alan Lyles. 2017. An in-home advanced robotic system to manage elderly home-care patients' medications: a pilot safety and usability study. *Clinical therapeutics*, 39, 5, 1054–1061.

[47] Annalyse Callahan Raub. 1981. Correlates of Computer Anxiety in College Students. (1981). Retrieved Feb. 14, 2022 from https://www.proquest.com/docview/303158028?fromopenview=true&pq-origsite=gscholar.

[48] Laurel D Riek. 2016. Robotics technology in mental health care. In *Artificial intelligence in behavioral and mental health care*. Elsevier, 185–203.

[49] Laurel D Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 245–246.

[50] Paul Robinette, Ayanna M Howard, and Alan R Wagner. 2017. Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, 47, 4, 425–436.

[51] [SW] Diligent Robotics, Moxi 2017.

[52] SoftBank Robotics. 2008. NAO the humanoid and programmable robot. (2008).

[53] Softbank Robotics. 2014. Pepper the humanoid and programmable robot. (2014).

[54] Selma Šabanović, Wan-Ling Chang, Casey C. Bennett, Jennifer A. Piatt, and David Hakken. 2015. A Robot of My Own: Participatory Design of Socially Assistive Robots for Independently Living Older Adults Diagnosed with Depression. en. In *Human Aspects of IT for the Aged Population. Design for Aging* (Lecture Notes in Computer Science). Jia Zhou and Gavriel Salvendy, (Eds.) Springer International Publishing, Cham, 104–114. ISBN: 978-3-319-20892-3. DOI: 10.1007/978-3-319-20892-3_11.

[55] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2121. (Mar. 2015), 1–8.

[56] Mariah L. Schrum, Michael Johnson, Muyleng Ghuy, and Matthew C. Gombolay. 2020. Four years in review: Statistical practices of likert scales in human-robot interaction studies. *ACM/IEEE International Conference on Human-Robot Interaction*, 43–52. ISBN: 9781450370578. DOI: 10.1145/3371382.3380739.

[57] Sarah Strohkorb Sebo, Priyanka Krishnamurthi, and Brian Scassellati. 2019. "i don't believe you": investigating the effects of robot trust violation and repair. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 57–65.

[58] Wayne Shelton. 2022. A review of "caregiving, carebots, and contagion". *Monash Bioethics Review*, 1–3.

[59] Adriana Tapus, Maja J. Mataric, and Brian Scassellati. 2007. Socially assistive robotics [Grand Challenges of Robotics]. *IEEE Robotics & Automation Magazine*, 14, 1, (Mar. 2007), 35–42. Conference Name: IEEE Robotics & Automation Magazine. DOI: 10.1109/MRA.2007.339605.

[60] Daniel Ullman and Bertram F. Malle. 2018. What Does it Mean to Trust a Robot? Steps Toward a Multidimensional Measure of Trust. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18). Association for Computing Machinery, New York, NY, USA, (Mar. 2018), 263–264. ISBN: 978-1-4503-5615-2. DOI: 10.1145/3173386.3176991.

[61] Alessandro Umbrico, Amedeo Cesta, Gabriella Cortellessa, and Andrea Orlandini. 2020. A Holistic Approach to Behavior Adaptation for Socially Assistive Robots. en. *International Journal of Social Robotics*, 12, 3, (July 2020), 617–637. DOI: 10.1007/s12369-019-00617-9.

[62] Aimee van Wynsberghe. 2013. Designing robots for care: care centered value-sensitive design. eng. *Science and Engineering Ethics*, 19, 2, (June 2013), 407–433. DOI: 10.1007/s11948-011-9343-6.

[63] Alessandra Vitanza, Grazia D'Onofrio, Francesco Ricciardi, Daniele Sancarlo, Antonio Greco, and Francesco Giuliani. 2019. Assistive Robots for the Elderly: Innovative Tools to Gather Health Relevant Data. en. In *Data Science for Healthcare: Methodologies and Applications*. Sergio Consoli, Diego Reforgiato Recupero, and Milan Petković, (Eds.) Springer International Publishing, Cham, 195–215. ISBN: 978-3-030-05249-2. DOI: 10.1007/978-3-030-05249-2_7.

[64] Rosalie H Wang, Aishwarya Sudhama, Momotaz Begum, Rajibul Huq, and Alex Mihailidis. 2017. Robots to assist daily activities: views of older adults with alzheimer's disease and their caregivers. *International psychogeriatrics*, 29, 1, 67–79.

[65] Rebecca Wiczorek and Dietrich Manzey. 2014. Supporting attention allocation in multitask environments: effects of likelihood alarm systems on trust, behavior, and performance. *Human factors*, 56, 7, 1209–1221.

[66] Rosemarie E. Yagoda and Douglas J. Gillan. 2012. You Want Me to Trust a ROBOT? The Development of a Human–Robot Interaction Trust Scale. en. *International Journal of Social Robotics*, 4, 3, (Aug. 2012), 235–248. DOI: 10.1007/s12369-012-0144-0.

[67] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. Evaluating effects of user experience and system transparency on trust in automation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 408–416.