

The Effects of Proactive Release Behaviors During Human-Robot Handovers

Zhao Han

Department of Computer Science
University of Massachusetts Lowell
Lowell, USA
zhan@cs.uml.edu

Holly Yanco

Department of Computer Science
University of Massachusetts Lowell
Lowell, USA
holly@cs.uml.edu

Abstract—Most research on human-robot handovers focuses on how the robot should approach human receivers and notify them of the readiness to take an object; few studies have investigated the effects of different release behaviors. Not releasing an object when a person desires to take it breaks handover fluency and creates a bad handover experience. In this paper, we investigate the effects of different release behaviors. Specifically, we study the benefits of a *proactive release*, during which the robot actively detects a human grasp effort pattern. In a 36-participant user study¹, results suggest proactive release is more efficient than rigid release (which only releases when the robot is fully stopped) and passive release (the robot detects pulling by checking if a threshold value is reached). Subjectively, the overall handover experience is improved: the proactive release is significantly better in terms of handover fluency and ease-of-taking.

Index Terms—Human-Robot Handover; Robots

I. INTRODUCTION

Handover is an essential step to achieve fluent human-robot collaboration [1]. However, it is not an easy task for a robot with manipulators to hand over an object to human fluently. The whole handover process consists of three phases: the approach phase, the signal phase, and the transfer phase [2]. The robot giver that possesses an object first approaches the human receiver who would like to get the object, signals the intent that the robot is ready to hand over the object, and transfers the object to the receiver.

Failures and bad user experiences can occur in any phase. Imagine you just went to your home office from the bedroom, and you asked your robot to bring the cell phone that you left behind. The robot approaches you and stands in front of your desk, behind computer monitors, waiting for you to stand up and take it. The experience would be better if the robot moved beside you so that you could get the object while sitting [3], [4]. Your robot also may fail to signal the readiness of releasing the object: i.e., the cell phone in the robot's hand is not pointed towards you but towards the robot. Intent is better signaled if the robot holds the cell phone horizontally and the side near you slightly upwards [5].

Compared to failures in approaching the receiver and signaling the readiness of the event, the transfer phase is the

¹The study is ready to reproduce with a Baxter robot. The code and environment setup is available at https://github.com/umhan35/handover_moveit

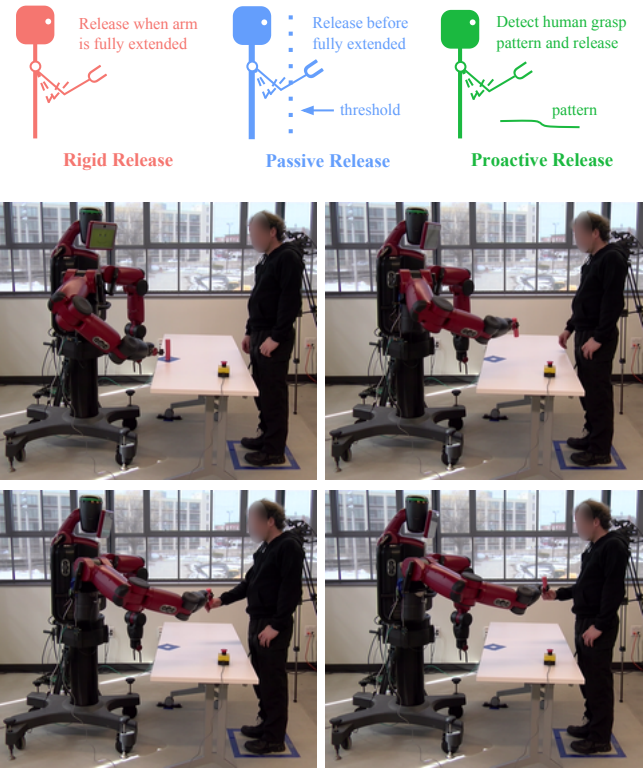


Fig. 1. To investigate the effects of the release behavior during human-robot handovers, we studied three release policies: proactive release, rigid release, and passive release. Participants took less time to complete handovers with the proactive release, which was rated as more fluent and easier to take the object. Overall, most participants preferred the proactive release.

most vital step because it determines whether the receiver will successfully retrieve the object; if it fails, the object might be dropped and broken [6], causing harm to humans. From the human's perspective, fluency [1] is hardly perceived: the receiver has to waste time waiting for the robot to release the object, and the experience of a certain number of failed attempts is frustrating [7]. To achieve fluent and comfortable handovers, robot designers need to tackle this problem.

To increase handover fluency and improve handover experience, we conducted a user study to investigate different ways for robots to release objects during a human-robot handover,

We study *proactive release* along with two other release policies, *rigid release* and *passive release* (Fig. 1). In the **rigid release policy**, the robot first fully extends its arm and, only when finished extending, detects pulling to release the object in the hand. During the **passive release policy**, the robot attempts to extend its arm fully while detecting pulling along the way, releasing the object accordingly. A pull is detected if the exerted force is over a predefined threshold. In the **proactive release policy**, the robot attempts to extend its arm fully while actively detecting a force change pattern inspired by human grasp effort along the way (e.g., Fig. 3), releasing the object early if a pull is detected.

In this paper, we present a within-subjects user study ($N = 36$) with three conditions to investigate the effects of different types of release behaviors in the handover process with a Rethink Robotics Baxter robot. We implemented existing mechanisms that attempt to achieve handover fluency in the human-robot interaction (HRI) literature, to allow the study to concentrate on the release behavior itself. Across a desk, participants attempt to take a foam cylinder from Baxter which stands one meter away [8], [9], shows gaze cues by looking at the object during the whole handover [10], and uses its right gripper [9] to grasp the object such that the bottom end of the cylinder is towards participants to easily grasp [5]. We also programmed Baxter so that the orientation of its arm and gripper is human-like, approaching from below to human arm height, rather than approaching from the top like an excavator, during the whole handover process (Fig. 2).

Results demonstrate handover efficiency and better user experience during the proactive release trials. Proactive release is also preferred over the other two release types by most participants. The proactive release is the most efficient among all three release types: participants complete the handover one second earlier than with the rigid and passive releases. Regarding handover fluency, the analysis of participants' Likert ratings reveals that proactive release is perceived to be more fluent than the rigid release. As expected, the analysis also suggests that it is easier to take the object during proactive release. In post-experiment surveys, participants noted that the robot "lets it go like a human" in the proactive release but "won't let it go" in the other two releases. However, given that the best existing mechanisms to achieve fluent handover experiences are implemented in all three releases, we did not find significant improvement in terms of trust, feeling comfortable, and capability. When we decompose the handover time, participants are neither early in preparing to take the object nor in attempting to get the object, but they did spend less time taking the object.

II. RELATED WORK

The literature in robotics and HRI has covered all three phases of the handover process, including the transfer phase, but none has investigated the effects of different kinds of release behaviors on improving fluency.

Through observations of a robot handing over a drink during a public demonstration, Cakmak *et al.* [7] confirmed that those

receivers who closely attend to the handover process tend to take the object early, while the robot is still moving or before prompting that the object is ready, and fail. The reason is attributed to the unclear signal phase and the ambiguous boundary between carrying and transfer. However, in the proposed solution [7], the robot fails to detect when the human desires to take the object. Chan *et al.* [6], [11] proposed design implications of handover controllers and conducted a human-human handover experiment. In the experiment, they discovered that both the giver and the receiver control their grip force according to load force, prompting the necessity of handling the grip force from a robotic giver. In the work done by Grigore *et al.* [12] to determine whether to release the object, the system monitors the human's focus of attention by analyzing whether the user simultaneously looks at and touches the cup through a head-mounted motion tracking system and a hand-worn glove. While the results show that adding a glove improves the handover success rate, it requires two extra pieces of intrusive equipment where, in our opinion, the glove can be replaced with a force sensor on the robot to improve the user experience. A few researchers [5], [13] have made use of the force sensor available on manipulators to detect object pull from the receiver, but, by the definition of pull, the receiver must draw with force to retrieve the object, in which scenario the robot passively releases the object, still breaking fluency and user experience. More recent work has been studying how to learn human preference [14] and model the approaching phase [15] during the handover process.

Because we also used Baxter's head-mounted display to express gaze behaviors, we surveyed the literature which investigated the effects of Baxter's virtual face or compared physical heads with virtual heads of different robots, including Baxter's; this work is discussed below in Section IV-A.

III. HYPOTHESES

We anticipate that the rigid release policy will have the worst user experience and delay the handover completion time because the robot refuses to release objects while extending its arm, even if the human receiver is ready to grasp the object. We expect that passive release will rank in second place because a forced pull to release is not as efficient as active detecting of force patterns. Thus, we propose the following hypotheses:

Hypothesis 1 (H1) - Release While Extending Arm. Detecting human grasp effort while extending its arm will improve the human receiver's experience of the handover, measured by early attempts, which affects completion time, and subjective measures.

Hypothesis 2 (H2) - Handover Efficiency. Proactive detection of human grasp effort makes the handover more efficient, measured by a reduction in handover completion time.

Hypothesis 3 (H3) - Handover Perception. The proactive release behavior increases the overall experience of the handover measured by subjective measures.

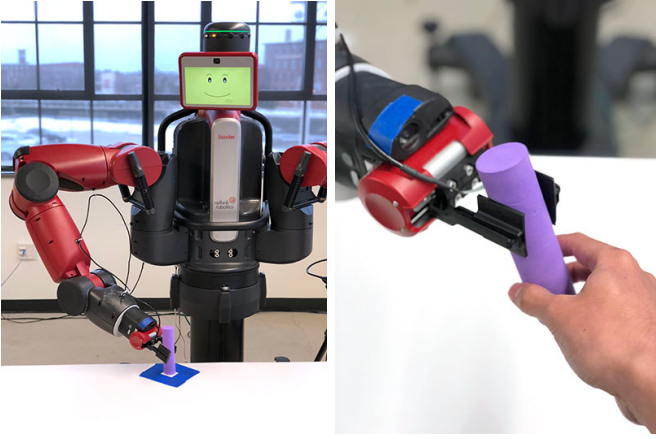


Fig. 2. Natural arm movement and the best mechanisms in the literature to achieve fluency are implemented in all three release policies. During the handover process, the robot, Baxter, turns its head and moves its eyes (face by Fitter *et al.*) [16] to keep looking at the cylinder [10]. The robot uses its right arm [9] to grasp the top part of the object, extending the bottom of the cylinder towards participants [5]. The arm of the robot is also as extended as possible to better signal its readiness [5].

IV. EXPERIMENT DESIGN

A. Robot Platform

A Baxter humanoid robot built by Rethink Robotics is used in this experiment. Baxter is equipped with two seven degree-of-freedom arms, each of which has a 2-finger gripper. Each gripper is equipped with a force sensor that outputs values ranging from 0 to 100, but these values are not available through Rethink’s API. We attached a thin, square force sensing resistor [17] to the right gripper; the sensor is connected to a voltage divider [18] to an interface board [19] to a computer. The standard screen on Baxter’s head is used to show the facial expression with gaze behaviors.

Some researchers have investigated Baxter’s facial display. Si *et al.* displayed a portrait photo of a human on Baxter’s facial screen and found that portraits on the screen “hurt the subjects’ trust and the perceived friendliness of the robot” [20]. Other researchers have been focused on displaying facial expressions only. Indeed, the original design of the display is to only show facial expressions. The paper on the process of developing Baxter by Rethink [21] explained that the display was designed to be pleasant and a lack of mouth can indicate that, as a colleague, Baxter will never retort. For the effect of a facial display versus a physical head capable of expressing emotions, Zhao *et al.* [22] compared the physical head of NAO and the facial display of Baxter, concluding that human observers understand both robots’ visual perspective at the same level. Sauppe and Mutlu [23] earlier found that eye movements following arm trajectories convey task status and intelligence for factory workers. In addition, Mizanoor *et al.* [24] found the same effect and that, when facial expressions are suitable, task performance is improved and the human partner is more engaged. To improve the validity of the experiment, we only used the facial expression designed by

Fitter *et al.* [16] with modifications to add eyeball movement to convey the handover task status and keep participants engaged.

B. The Procedure and The Task

Each participant follows a within-subjects experiment design, to allow participants to compare the three different release policies by the robot. The order of the three rounds, one for each of the types of release behaviors, is randomized prior to each participant to control for order effects through full counterbalancing. We used a script of instructions to control the variance that could introduced by instruction difference.

Before the experiment, participants were asked to read and sign an informed consent form which let participants know the purpose of the study, the duration, and the whole procedure. We also asked participants to ask clarification questions to ensure their understanding of the material. This study was approved by the ethics committee at the authors’ institution.

Participants were asked to stand inside of a square of blue tape on the floor. The robot handed over the foam cylinder from a blue square on the desk. After participants took the object from the robot, they were asked to put the object back in the blue square after the robot retracts its arm because the participant’s attention needs to be redirected to avoid potential judgment over the robot, also known as washout, between handovers, as in [5], [7], [10]. Participants were also told that there were would be three sets of handovers, each with ten trials (designed to avoid participants actively looking for the differences in different handover release behaviors, similar to [25]). Ten trials for each round allowed participants to build a consistent feeling and opinion on each type of release behavior, making the answers to the questionnaire also more consistent and thus reliable. At the beginning of each set, participants were reminded to think aloud, to say whatever comes into their mind. At the end of each set, the participants were asked to sit down to fill out part of the questionnaire, then the remainder of it at the end of the experiment.

C. The Handover Process for the Robot

Based on prior human-human handover studies on how a giver should approach a receiver [8], [9], the robot is approximately one meter away from participants. In accordance with gazing behavior experiments in the literature, Baxter turns its head and moves its eyeballs to look at the cylinder during the whole handover process [10] to avoid shifting the human receiver’s attention, which negatively affects the handover efficiency [13]. As suggested by Koay *et al.* [9], Baxter’s right arm and gripper are used. Inspired by studies of human preference of object configuration [5], the robot grasps the top of the cylinder, extending the bottom towards participants around 10 cm so that the object’s orientation will not affect participants’ feeling of how easy it is to take it. The robot’s arm is also as extended as possible to better signal the readiness to transfer the object [5]. At the end of each trial, the robot turns its head and moves its eyeballs to look at the participant’s head. Baxter then hands over the cylinder and releases it according to the three policies, depending upon the round. When released,

Baxter retracts its arm. Fig. 2 shows a glimpse of the gazing behavior, head movement, and the object configuration.

D. Policy Implementation

Because the experiment is exploring the effects of the three different release behaviors, all other movements executed by Baxter are preconfigured, including head movements, facial expressions, gaze directions, approaching and grasping the object, extending its right arm, approaching participants, and retracting its arm. In addition, because the built-in inverse kinematics solver always generates trajectories that approach objects from above, like an excavator, we used the MoveIt! motion planning framework [26] to specify the orientation of the gripper so that Baxter approaches the object and participants from the side in a human-like manner, shown in the right image of Fig. 2. To ensure repeatability, Baxter’s motion remains the same across all conditions.

The rigid release policy serves as a baseline. The desk between Baxter and participants is modeled as a collision object so that Baxter does not hit participants. In the rigid release, the robot is programmed such that force detection only starts when the specified trajectory is completely executed. A voltage value, proportional to the force applied to the surface of the force sensing resistor, is received every 256 ms, the default rate in the hardware API [27]. We empirically determined the handover force change using a range of grasping attempts from a light grasp to a heavy pull and set its corresponding voltage change to 0.03 to avoid accidental drops caused by arm movement and unstable stoppage.

The only difference between the passive release policy and the rigid release policy is that the force detection in the passive release policy starts when the robot’s arm passes a distance threshold to be within a participant’s reach.

In the proactive release policy, the force detection starts at the same time and distance as in the passive release policy; however, it receives a message every 1 ms so that we are able to look for a decreasing force pattern presented during grasping. Fig. 3 shows an example of the pattern that we found after observing the force data stream during a range of grasping attempts in varying degrees, from a light grasp to a heavy pull. The voltage values decrease because the friction force, which is not orthogonal to the pressing force, counteracts the pressing force. In the implementation, we used a moving average technique to calculate the average values of windows of 180 voltage data points, collected over 180 ms. The program then checks the past 90 windows to determine if 35% of the average values are decreasing from window to window to determine whether to release. Similar to the voltage threshold used in the rigid and passive release policies, this threshold is selected to avoid accidental drops caused by arm movement and unstable stoppage.

E. Data Collection and Measures

All experiments were videotaped then coded to extract timing and frequency data. Additionally, the task completion time was logged by the software from when Baxter grasps

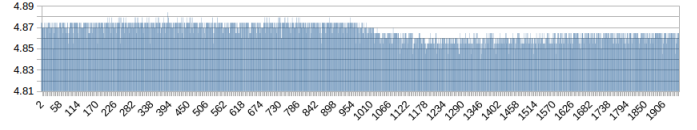


Fig. 3. One example of the decreasing pattern presented during human grasping. The y-axis represents the voltage proportional to force and the x-axis represents time in seconds. Voltage values decrease during grasping because the friction force, which is not orthogonal to the pressing force, counteracts the pressing force.

the object (t_h) to when the participant successfully takes the object or Baxter releases the object (t_r).

We extracted four pieces of timing and frequency information from the videos: when participants prepare to take the cylinder (t_p) by starting to move their arm or hand to take it; whether Baxter stops moving without participants touching the object (f_s); when participants touched the object (t_t), meaning there is no space between the object and participants’ fingers; and if Baxter accidentally dropped the object (f_d).

In order to gather the timing data, we set up a total of four camcorders. One camcorder (C_4) was placed above the scene, allowing coders to code t_p , t_t and t_r . Another camcorder (C_1) was placed to the left side of participants, which helped to code f_s and disambiguate t_t . The remaining two camcorders were placed to the right side of participants: one behind the left arm of Baxter (C_2) to verify t_p and record the facial expression of participants for subjective analysis, and another to the right rear of participants (C_3) to code t_h .

With the data, we are able to explore some research questions to support our hypotheses. $t_p - t_h$, or simply t_p , answers how early people attempted to take the object. $t_r - t_t$ answers how long and how easy it is for people to take the object. If t_t is later than when the robot stopped, we count it as one early handover attempt, i.e., f_s . The total number of appearances of f_d shows dropped cases in different release policies.

Participants completed a questionnaire throughout the study to capture subjective experience. Participants first answered demographic questions. After each round, participants were asked to complete a page of the questionnaire with the same Likert scale questions regarding fluency, ease-of-taking, trust, comfort, and capability, listed in Table I. The question about the fluency aspect is inspired by Hoffman [1]. After each Likert question, participants were encouraged to leave comments. When they completed the questions for the last set, they were encouraged to change their previous answers, if desired, after a comprehensive comparison of all three release behaviors. The post-session questions also helped participants avoid fatigue that might affect study results. After the last round, they were asked to answer the remainder of the questionnaire which consisted of free-form questions regarding the different types of release behaviors and the overall handover experience.

The experimenter and one independent coder coded the videos of 36 participants frame by frame for the timing information of t_h , t_p , t_t and t_r , and the frequency of f_s , as detailed earlier. The coder and experimenter jointly coded a

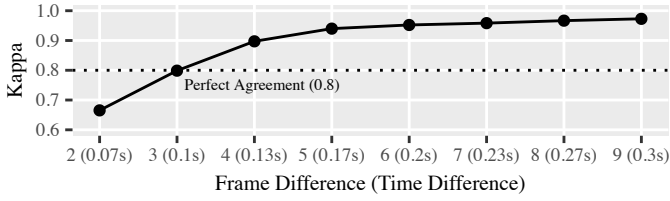


Fig. 4. The κ values for video coding to illustrate almost perfect intercoder agreement [28] when the frame difference is increased to 3 ($p < 0.001$ across all frame differences).

random of 10% of the videos; the remainder of the videos were coded solely by the experimenter. Due to the accuracy achieved through frame-by-frame coding, Cohen's κ shows strong agreement between the two coders. Because the videos are shot at 30 frames per second (FPS), the agreement of one event happening at the same time depends on the allowable frame difference chosen. Shown in Fig. 4, we achieve a κ value over 0.8 (almost perfect agreement [28]; $p < 0.001$) when the frame difference is 3, in which the time difference is only 0.1 second. When the frame difference is increased to 4 (0.13 second), κ is increased to 0.9 ($p < 0.001$).

F. Participants

We recruited 41 participants from the university community and the surrounding city using flyers and email lists. Five were excluded due to hardware failures. To control for order effects, we continued recruiting until we had valid data from 36 participants (6 multiples of 3! = 6). For the 36 participants whose data was used, from the answers to how they heard about us, we estimate that 14 (39%) were non-students and 22 (61%) were students. Participants were given a \$15 gift card. Ages ranged from 18 to 57 ($M = 29$, $SD = 12$), with 21 male and 15 female. Three said they were left-handed, while the rest were right-handed. When asked if they had experience with robots, 8 agreed, 26 disagreed and 2 chose neutral.

V. RESULTS

We used R to analyze the data logged by the software and coded from the videos. M used without standard deviation values denotes median values throughout this section.

A. Preference

Twenty-nine (of 36) participants explicitly stated a single preference in the questionnaire after the experiment. Twenty of these 29 (69%) participants preferred the proactive release, 7 (24%) liked the passive release more, and 2 selected the rigid release, as shown in Fig. 5. For the remaining 7 participants, 2 chose passive/rigid because they preferred a forced pull, 3 did not notice any difference, and the other 2 did not answer the question explicitly or imply any preference.

We performed a multinomial goodness-of-fit test on the preference data with the sample size of 29 and found a statistically significant result ($p < 0.001$). We also performed post-hoc binomial tests with Holm-Bonferroni correction [29] for pairwise comparisons, which shows statistically significant

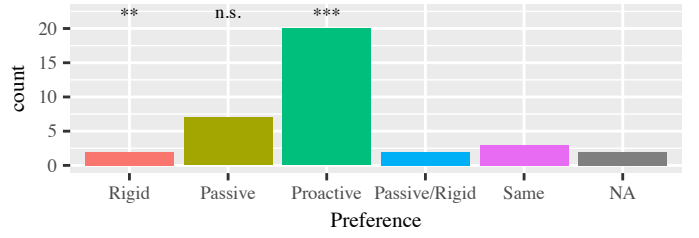


Fig. 5. Most participants preferred the proactive release. A multinomial goodness-of-fit test and post-hoc comparisons show statistically significant differences in the rigid and proactive releases.

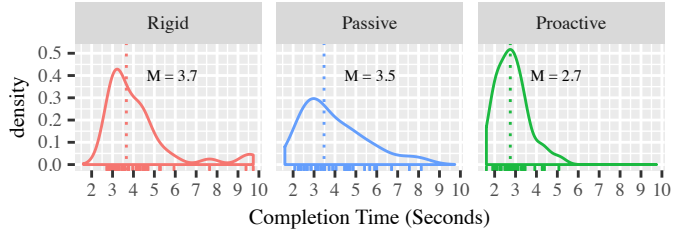


Fig. 6. The density plots with rug plots at the bottom and median lines show the highly skewed distributions of the handover completion time for rigid and passive release types and a slight departure from the normal distribution for the proactive release. Along with the violation of homoscedasticity, we used Friedman's test [30] for handover timing analysis.

preferences in rigid ($p < 0.01$) and proactive ($p < 0.001$) but not in passive ($n.s.$).

B. Objective Measures: Handover Completion Time

We first analyzed the overall handover completion time, which supports H2: the proactive release behavior makes the handover process more efficient. Of the 3 conditions, the rigid and passive releases are both time-consuming: the former takes $M = 3.7$ seconds and the latter takes $M = 3.5$ seconds, between which we did not find a statistically significant difference. Unsurprisingly, we found strong statistically significant differences between the proactive trials and the rigid trials ($M = 2.7$ vs. $M = 3.7$, $p < 0.0001$) and between proactive and passive ($M = 2.7$ vs. $M = 3.5$, $p < 0.0001$). There is no statistically detectable difference between rigid and passive, but we argue the effect at the end of this subsection by taking a closer look at the density plots shown in Fig. 6.

To understand whether there is a difference in handover completion time with different types of release behaviors, we first considered performing a one-way repeated measures ANOVA. However, both the assumption of normality and homoscedasticity are violated, at a level that ANOVA is not robust anymore. Shown in Fig. 6, we found that the handover completion time data are not normally distributed in each release type but rather highly right-skewed except for the proactive type. The non-normality is confirmed by the Shapiro-Wilk normality tests (extreme violation in rigid data: $W = 0.73$, $p < 0.0001$, more than moderate violation in passive data: $W = 0.90$, $p < 0.001$, marginal violation in proactive data: $W = 0.94$, $p = 0.051$). Indeed, one-sample Kolmogorov-Smirnov goodness-of-fit test in each release type

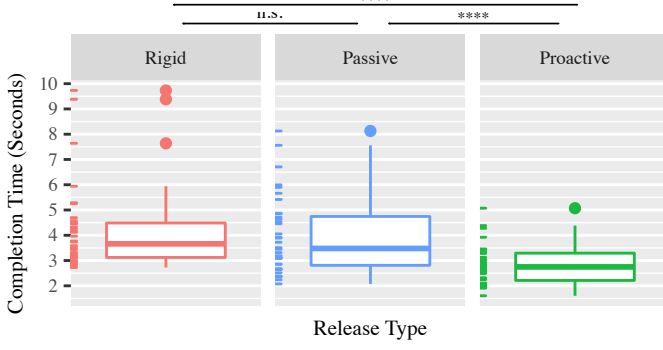


Fig. 7. The box plots with rug plots illustrating the completion time, with results from Friedman’s test reports statistical significance and the post-hoc Wilcoxon signed-rank test [31] results.

shows that the completion time follows a log-normal distribution (rigid: $D = 0.13, n.s.$, passive: $D = 0.11, n.s.$, proactive: $D = 0.10, n.s.$). The assumption of homoscedasticity had also been violated, suggested by the Brown-Forsythe test ($F(2, 105) = 3.31, p < 0.05$).

In the past, ANOVA and the t test for the post-hoc pairwise comparisons have been shown to be relatively insensitive to moderate normality violation for the size of each sample over 30 [32], [33] and relatively insensitive to the homogeneity of variance assumption violation. But because both the normality violations in rigid and passive are more than moderate, we conduct the non-parametric asymptotic Friedman’s test [30].

The Friedman’s test shows that there is a statistically significant effect of release types on the median handover completion time at the $p < .05$ level for the three conditions ($\chi^2(2) = 34.06, p < 0.0001$). As illustrated on the top of the box plots in Fig. 7, we performed post-hoc pairwise comparisons using Wilcoxon signed-rank tests [31] with Holm-Bonferroni correction, which show a statistically significant difference ($p < 0.0001$) between rigid ($M = 3.7$) and proactive ($M = 2.7$), and a significant difference ($p < 0.0001$) between passive ($M = 3.5$) and proactive ($M = 2.7$). This statistical result confirms that the decreasing pattern detection implemented in the proactive release policy is effective.

The Wilcoxon signed-rank test did not detect a significant difference between rigid and passive. However, from the left two density plots in Fig. 6, comparing the left skew in each density curve tells us that there are a number of people who are able to complete the handover in 2 to 2.5 seconds during the passive trials, which did not happen in the rigid trials.

There are no order effects found by examining the handover completion time across different orders. The completion time in any order does not follow normal distribution, confirmed by Shapiro Test for order 1 data ($p < 0.0001$), order 2 data ($p < 0.0001$), and order 3 data ($p < 0.001$). However, the ANOVA assumption of homoscedasticity is confirmed by the Brown-Forsythe test ($F(2, 105) = 0.28, n.s.$). Because of the extreme violation of the normality violation, we conducted Friedman’s test ($\chi^2(2) = 2.67, n.s.$).

TABLE I
LIKERT SCALE QUESTIONNAIRE (Cronbach’s $\alpha = 0.77$)

Fluency (Cronbach’s $\alpha = 0.68$ if dropped)
<i>The robot contributed to the fluency of the handover: the robot handed over like a human.</i>
Ease-of-taking (Cronbach’s $\alpha = 0.72$ if dropped)
<i>It is easy to take the object from the robot.</i>
Trust (Cronbach’s $\alpha = 0.70$ if dropped)
<i>I trust the robot to do the right thing at the right time.</i>
Discomfort (Cronbach’s $\alpha = 0.78$ if dropped)
<i>I feel uncomfortable with the robot.</i>
Capability (Cronbach’s $\alpha = 0.72$ if dropped)
<i>The robot was capable of handing over the object.</i>

* Likert items are coded as -3 (Strongly Disagree), -2 (Disagree), -1 (Moderately Disagree), 0 (Neutral), 1 (Moderately Agree), 2 (Agree), and 3 (Strongly Agree).

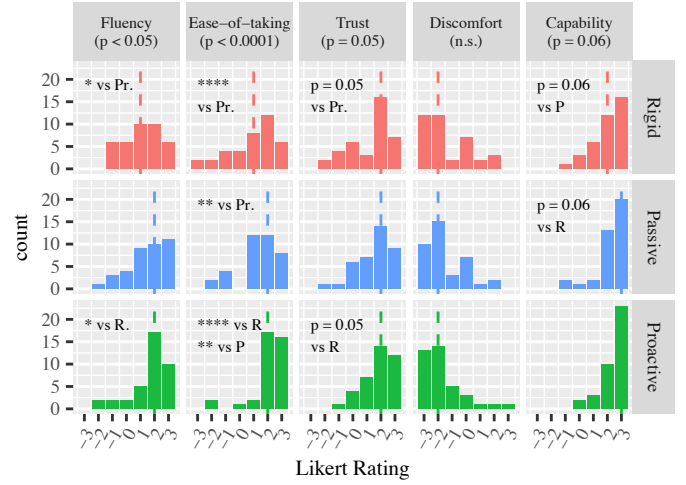


Fig. 8. All raw questionnaire responses. After conducting Friedmans tests and Wilcoxon signed-rank tests, we found statistically significant results in responses to the fluency and ease-of-taking questions. Trends are found in responses to the trust and capability questions.

C. Subjective Measures: Questionnaire Responses

Table I lists all Likert questions from the questionnaire and the codes for each Likert item. Indicated by Fig. 8, all of the data significantly departs from the normal distribution for each question-release type. Because Likert ratings are nominal data, we conducted Friedman’s test again. Fig. 8 plots all Likert ratings and shows the statistical results and the median values. Fig. 9 summarizes the data using box plots.

After reversing the ratings of responses about discomfort to comfort, Cronbach’s alpha reports an acceptable level of internal consistency reliability ($\alpha = 0.77$) [34]. The Cronbach’s α values when a question is dropped are also given in Table I: 0.68 for fluency, 0.72 for ease-of-taking, 0.70 for trust, 0.78 for discomfort, and 0.72 for capability.

In the fluency question responses, Friedman’s test suggests there is a statistically significant difference ($\chi^2(2) = 21.46, p < 0.05$). Post-hoc pairwise comparisons using Wilcoxon signed-rank tests with Holm-Bonferroni correction shows that there is a statistically significant difference between the rigid ($M = 1$) and the proactive ($M = 2$) release behaviors

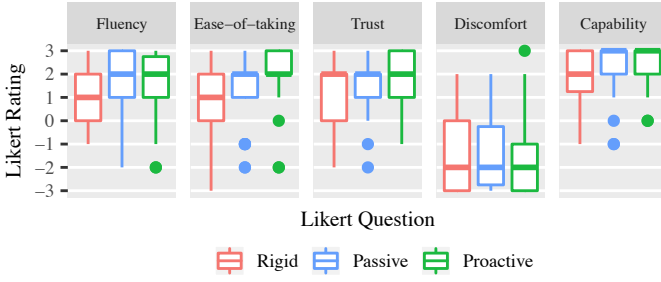


Fig. 9. Box plots summarize the questionnaire responses.

($p < 0.05$). However, there are no statistically detectable differences between rigid-passive and proactive-passive pairs.

In the ease-of-taking Likert ratings, Friedman's test suggests there is a stronger statistically significant difference ($\chi^2(2) = 7.66, p < 0.0001$). The same type of post-hoc pairwise comparisons indicate that the proactive release ($M = 2$) is significantly better than the other two (rigid - $M = 1$: $p < 0.0001$; passive - $M = 2$: $p < 0.01$). Unsurprisingly, there is no detectable difference between rigid and passive.

However, we only found a marginally statistically significant difference in the trust ratings through Friedman's test ($\chi^2(2) = 5.86, p = 0.05$). Post-hoc pairwise comparisons using Wilcoxon signed-rank tests with Holm-Bonferroni correction show that the trend is between rigid and proactive ($p = 0.05$). For responses regarding discomfort, we did not find a statistically significant difference ($\chi^2(2) = 2.23, n.s.$). Friedman's test suggests there is only a trend in the capability between three release behaviors ($\chi^2(2) = 5.64, p = 0.06$). Post-hoc pairwise comparisons suggest the trend in between rigid and passive ($p = 0.06$). The release type is not a significant factor that affects whether the robotic handover is comfortable to humans and its handover capability.

We also did not find any order effects via Friedman's test on each question response data across different orders ($n.s.$).

D. Breakdown of Handover Completion Time

Across three release policies, we did not find statistically significant effects on when participants prepare to take the object (t_p), whether participants attempted to get the object while the robot was still extending its arm (f_s), or the number of drops (f_d). However, participants spent half of the time taking the object ($M = 0.507$) in the proactive trials, consistent with the finding in the completion time.

Similar to the timing data of handover completion time, we chose to perform Friedman's test for all the analysis in this section. Despite the confirmation of homogeneity by the Brown-Forsythe test (t_p : $F(2, 105) = 0.89, n.s.$; t_r : $F(2, 105) = 0.04, n.s.$; release duration: $F(2, 105) = 0.45, n.s.$), the violation of normality is extreme across all handover timing data after the Shapiro-Wilk normality tests are performed across all release types (t_p : $W = 0.91, p < 0.01$ vs. $W = 0.91, p < 0.01$ vs. $W = 0.88, p < 0.01$; t_r : $W = 0.73, p < 0.0001$ vs. $W = 0.90, p < 0.001$ vs. $W = 0.90, p < 0.001$; release

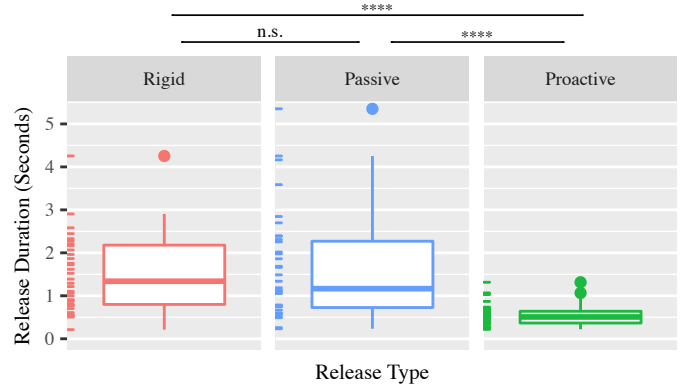


Fig. 10. The box plots with rug plots at the bottom show the release duration. Friedman's test reports statistical significance and the post-hoc Wilcoxon signed-rank test [31] results are also shown. Note that, in the rigid box plot, two outliers between 6 to 8 are not shown.

duration: $W = 0.73, p < 0.0001$ vs. $W = 0.87, p < 0.0001$ vs. $W = 0.90, p < 0.001$).

There are also no order effects found. Given the extreme violation of normality on all orders ($W = 0.92, p < 0.01$ vs. $W = 0.90, p < 0.01$ vs. $W = 0.86, p < 0.001$), we conducted Friedman's test (t_p : $\chi^2(2) = 0.22, n.s.$; t_t : $\chi^2(2) = 0.06, n.s.$; release duration: $\chi^2(2) = 0.72, n.s.$).

By analyzing when participants prepare to take the cylinder, t_p , we found that participants did not attempt to move their arms or hands earlier to take the object in any release condition. This is expected because release does not affect how human receivers approach the object. There is no detectable difference across the three release types ($M = 0.673$ vs. $M = 0.605$ vs. $M = 0.648$), $\chi^2(2) = 3.17, n.s.$.

We also did not find a statistical difference in whether participants attempted to take the object early while the robot was still extending its arm: in all release types, half of the participants did while the other half did not. We performed multinomial goodness-of-fit tests on the f_s frequency data across all release types and found no significant results.

As expected, by analyzing the f_d frequency data using a multinomial goodness-of-fit test, there is no detectable difference in the number of drops across the three conditions. In the 1080 trials, there were 2, 4, and 8 drops in the rigid, passive, and proactive conditions, respectively. All of the drops were due to participants, e.g., focusing on the head while grasping.

When analyzing when participants touched the object, t_t , we found significant differences (Friedman's test: $\chi^2(2) = 36.17, p < 0.0001$) between proactive and rigid ($p < 0.0001$), between proactive and passive ($p < 0.0001$), but not between rigid and passive ($n.s.$), reported by the post-hoc Wilcoxon signed-rank tests. In proactive trials ($M = 1.884$), participants touched the cylinder 88 ms earlier than rigid ($M = 1.972$) but 46 ms later than passive ($M = 1.838$). This may suggest release behaviors does not affect early attempts.

Finally we analyzed the release duration, calculated by $t_r - t_t$ (Fig. 10). Results show participants only spend half of the time in proactive trials ($M = 0.507$) than in rigid ($M = 1.339$)

and passive ($M = 1.166$). Similar to t_t , there are significant differences ($\chi^2(2) = 41.17, p < 0.0001$) between proactive and rigid ($p < 0.0001$), between proactive and passive ($p < 0.0001$), but not between rigid and passive (*n.s.*). This result is consistent with the finding in the handover completion time.

VI. DISCUSSION

We did not find consistent evidence to support H1 that releasing while extending the robot's arm increases the human receiver's experience. Against H1, there was no single release policy where more participants had early attempts. There were also no detectable differences in completion time between rigid and proactive releases. However, participants rated the passive release as easier to take ($M = \textit{Agree}$) than the rigid release ($M = \textit{Moderately Agree}$).

H2 is strongly supported: proactive detection of the human grasp made the handover more efficient. There is a statistically significant difference in completion time between the proactive policy ($M = 2.7$) and both rigid ($M = 3.7$) and passive ($M = 3.5$). There is a statistically significant difference in release duration between the proactive policy ($M = 0.507$) and both rigid ($M = 1.339$) and passive ($M = 1.166$).

H3 also holds: the proactive release behavior improved the overall experience of the handover process. In questionnaire responses, participants rated proactive being more fluent ($M = \textit{Agree}$) than rigid ($M = \textit{Moderately Agree}$), a statistically significant difference ($p < 0.05$). The results also suggest that it is easier ($p < 0.0001$) to take the object from the robot during proactive trials ($M = \textit{Agree}$) than rigid and passive ($M = \textit{Moderately Agree}$; $p < 0.01$ and $p < 0.0001$, respectively).

When trying to take the object from the robot, we observed that most participants tended to hold or touch the object without pulling. Twenty-five (70%) participants complained multiple times about the rigid release seeming inconsistent, finding it sometimes easy and often hard (9 complained only during think-aloud; 2 only on the questionnaire; and 14 in both). Twenty-seven (75%) participants had the same issue with passive release (7 complained only during think-aloud, 2 only on the questionnaire, and 18 in both). We attribute this perception to the different ways that participants would grasp the object in different trials, with the robot not reacting any differently in either of those two release strategies.

With the proactive release, 12 (33%) participants expressed just once or twice that they experienced a little pull or a slight delay (7 only during think-aloud, 3 only on the questionnaire, and 2 in both). This difference in how people experience the consistency of the release policy explains why the rigid and passive releases are less efficient, since it takes more time to react to seemingly inconsistent releases, supporting H2.

Rigid and passive releases also had less favorable ratings in ease-of-taking than the proactive release, supporting H3. The common observation of inconsistency between the rigid and passive releases also explains why there is no detectable difference in ease-of-taking ratings between them.

For fluency, thirteen (36%) participants explicitly described the proactive release as smooth or fluid on the questionnaire

while only 2 participants described passive release this way and none did for rigid. Ten (28%) participants and eleven (30%) participants expressed experiencing delay or pull with the rigid and passive releases, respectively.

There is a potential confound with the sensor sampling rate, which is 256 ms for rigid and passive but 1 ms for proactive, as early detection may cause higher ease-to-take perception ratings and shorter completion time. However, because all policy implementations start receiving sensor data when the program starts before the experiment, the faster sampling rate does not always give the proactive release an advantage of early detection. In theory, the proactive release only requires 32 decreasing samples ($90 \text{ windows} \times 35\%$, assuming all of the last windows have the values) to identify a grasp while the other two policies require a single sample at the 256 ms rate that is above the threshold. Depending upon the timing of the participants grasp, it could actually be the case that a single data point detected at the 256 ms rate would be registered before a grasp was detected in the proactive policy, meaning that 12.5% of the time, the single data point method in the rigid and passive releases would be faster to record the grasp.

A grasp is not only physically touching an object, but also feeling these events, in which a 200 ms neuromuscular response time exists [35], which delays the acknowledgement of not releasing in the brain. During the experiment, we did not observe that the longer sampling rate had much of an impact, perhaps given that it is just 56 ms longer than the 200 ms neuromuscular response time in addition to the randomness discussed in the paragraph above. Rather, it is the early attempt, in which the robot does not release during rigid trials, or a light pull or simply holding without pulling, in which the threshold value is not met in both the rigid and passive trials, that affects participants perception of ease-to-take and delays handover completion and release time.

However, at this time, the difference in sampling rates could be a confound, and this paper has only shown that the combination of proactive release and its sampling rate is what makes a difference over rigid and passive releases with their sampling rates. A future study could be conducted with identical sampling rates, although we posit that would then favor the rigid and passive releases (1 ms sampling) over the proactive release (32 ms minimum). Ultimately, it would be difficult to guarantee that a persons touch was registered in an identical amount of time, regardless of the release policy.

VII. CONCLUSION

We investigated three different release behaviors for human-robot handovers with 36 participants. Results with strong statistical evidence suggest that proactive release, in which the robot actively detects a human grasp effort pattern, significantly improves the handover experience and efficiency.

ACKNOWLEDGEMENTS

This work has been supported in part by the National Science Foundation (IIS-1426968 and IIS-1763469) and the Office of Naval Research (N00014-18-1-2503)

REFERENCES

- [1] G. Hoffman, "Evaluating fluency in human-robot collaboration," in *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013.
- [2] K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli, "Towards seamless human-robot handovers," *Journal of Human-Robot Interaction (JHRI)*, vol. 2, no. 1, pp. 112–132, 2013.
- [3] J. Mainprice, E. A. Sisbot, T. Simeon, and R. Alami, "Planning safe and legible hand-over motions for human-robot interaction," in *IARP Workshop on Technical Challenges for Dependable Robots in Human Environments*, 2010.
- [4] J. Mainprice, M. Gharbi, T. Simeon, and R. Alami, "Sharing effort in planning human-robot handover tasks," in *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2012, pp. 764–770.
- [5] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, "Human preferences for robot-human hand-over configurations," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 1986–1993.
- [6] W. P. Chan, C. A. Parker, H. M. V. der Loos, and E. A. Croft, "A human-inspired object handover controller," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 971–983, 2013.
- [7] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human hand-overs," in *Proceedings of the 6th international conference on Human-robot interaction (HRI)*, 2011, pp. 489–496.
- [8] P. Basili, M. Huber, T. Brandt, S. Hirche, and S. Glasauer, "Investigating human-human approach and hand-over," in *Human centered robot systems*, 2009, pp. 151–160.
- [9] K. L. Koay, E. A. Sisbot, D. S. Syrdal, M. L. Walters, K. Dautenhahn, and R. Alami, "Exploratory study of a robot approaching a person in the context of handing over an object," in *Multidisciplinary Collaboration for Socially Assistive Robotics*, 2007, pp. 18–24.
- [10] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI)*, 2014, pp. 334–341.
- [11] W. P. Chan, C. A. Parker, H. M. V. der Loos, and E. A. Croft, "Grip forces and load forces in handovers: implications for designing human-robot handover controllers," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (HRI)*, 2012, pp. 9–16.
- [12] E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards, "Joint action understanding improves robot-to-human object handover," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 4622–4629.
- [13] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI)*, 2014, pp. 49–56.
- [14] A. C. H. Quispe, E. Martinson, and K. Oguchi, "Learning user preferences for robot-human handovers," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 834–839.
- [15] S. Parastegari, B. Abbasi, E. Noohi, and M. Zefran, "Modeling human reaching phase in human-human object handover with application in robot-human handover," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3597–3602.
- [16] N. T. Fitter and K. J. Kuchenbecker, "Designing and assessing expressive open-source faces for the baxter robot," in *International Conference on Social Robotics*, 2016, pp. 340–350.
- [17] "Interlink Electronics 1.5" Square 20N FSR," <https://www.phidgets.com/?tier=3&catid=6&pcid=4&prodid=209>, accessed: 2018-11-8.
- [18] "Voltage Divider," <https://www.phidgets.com/?tier=3&catid=49&pcid=42&prodid=92>, accessed: 2018-11-8.
- [19] "PhidgetInterfaceKit 8/8/8," <https://www.phidgets.com/?tier=3&catid=2&pcid=1&prodid=18>, accessed: 2018-11-8.
- [20] M. Si and J. D. McDaniel, "Establish trust and express attitude for a non-humanoid robot," in *Conference of the Cognitive Science Society*, 2016.
- [21] C. Fitzgerald, "Developing baxter," in *Proceedings of the 5th IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, 2013, pp. 1–6.
- [22] X. Zhao, C. Cusimano, and B. F. Malle, "Do people spontaneously take a robot's visual perspective?" in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 335–342.
- [23] A. Sauppé and B. Mutlu, "The social impact of a robot co-worker in industrial settings," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems (CHI)*, 2015, pp. 3613–3622.
- [24] R. S. Mizanoor, D. A. Spencer, X. Wang, and Y. Wang, "Dynamic emotion-based human-robot collaborative assembly in manufacturing: The preliminary concepts," in *Workshop on Human-Robot Collaboration for Industrial Manufacturing at RSS '14*, 2014.
- [25] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 51–58.
- [26] "MoveIt! Motion Planning Framework," <https://moveit.ros.org/>, accessed: 2018-11-8.
- [27] "Phidgets C API," http://wiki.ros.org/phidgets_api?distro=indigo, accessed: 2018-11-8.
- [28] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.
- [29] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, pp. 65–70, 1979.
- [30] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the american statistical association*, vol. 32, no. 200, pp. 675–701, 1937.
- [31] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.
- [32] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," *Review of Educational Research*, vol. 42, no. 3, pp. 237–288, 1972.
- [33] R. R. Pagano, *Understanding statistics in the behavioral sciences*. Cengage Learning, 2012.
- [34] R. F. DeVellis, *Scale Development: Theory and applications*. Sage publications, 2016, vol. 26.
- [35] M. Minsky, O.-y. Ming, O. Steele, F. P. Brooks Jr, and M. Behensky, "Feeling and seeing: issues in force display," in *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, vol. 24, no. 2. ACM, 1990, pp. 235–241.