

Do Results in Experiments with Virtual Robots in Augmented Reality Transfer To Physical Robots? An Experiment Design

Xiangfei Kong

RARE Lab

Department of Computer Science and Engineering
University of South Florida
Tampa, Florida, USA
xiangfeikong@usf.edu

Zhao Han

RARE Lab

Department of Computer Science and Engineering
University of South Florida
Tampa, Florida, USA
zhaohan@usf.edu

ABSTRACT

Entry to human-robot interaction research, e.g., conducting empirical experiments, faces a significant economic barrier due to the high cost of physical robots, ranging from thousands to tens of thousands, if not millions. This cost issue also severely limits the field's ability to replicate user studies and reproduce the results to verify their reliability, thus offering more confidence to incorporate these findings. Although virtual reality (VR) user studies present a potential solution, it is unclear whether we can confidently transfer the findings to physical robots and physical environments because VR isolates both the physical robot and the physical world where robots operate. To address this issue, we propose to leverage augmented reality (AR) only to simulate a virtual robot but retain the physical environment. Specifically, we designed an experiment to compare virtual and physical robots in a physical mobile manipulation task environment, involving manipulation and navigation tasks for generalizability. To further improve ecological validity, we evaluate both types of robots subjectively regarding usability, trust, and personal preference, which HRI research has shown to be widely used in HRI. The results of this work will benefit many researchers studying these important issues in robotics. If virtual robots mixed into the physical world are not worse than physical robots, it opens new possibilities for empirical research with cost-free virtual robots and the results are transferable to physical situations.

CCS CONCEPTS

• **Computer systems organization** → **Robotics**; • **Human-centered computing** → **Mixed / augmented reality**; **Interaction design process and methods**; • **General and reference** → **Reliability**.

KEYWORDS

Human-robot interaction (HRI). Replication. Reproducibility. Augmented reality (AR). Performance

ACM Reference Format:

Xiangfei Kong and Zhao Han. 2024. Do Results in Experiments with Virtual Robots in Augmented Reality Transfer To Physical Robots? An Experiment Design. In *Proceedings of the 4th Workshop YOUR Study Design! Participatory*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WYSD '24, March 11, 2024, Boulder, CO, USA

© 2024 Copyright held by the owner/author(s).

critique and refinement of human-robot interaction user studies (WYSD '24), March 11, 2024, Boulder, CO, USA. 4 pages.

1 INTRODUCTION

Human-robot interaction (HRI) leverages a wide range of academic disciplines and various research methods [17], the former including social psychology, robot design, and the latter including quantitative and qualitative methods in settings like labs and the wild. Recently, there has been increasing research on reproducibility and replicability [8, 12–14, 21, 27, 28]. The goal is to improve confidence and trust in the findings of user studies as well as for researchers to easily share and expand upon each other's findings, thus facilitating rapid advancement in the field.

Yet, the cost of physical robots presents an economic barrier that limits the replication and reproduction of HRI experiments with physical robots. For instance, the starting price for a two-link education robot arm [5] with a minimally useful reach (45cm) is already about \$1,700. For mobile manipulators like the Hello Robot Stretch [3] and Fetch [2], frequently used in HRI studies, range from about \$25,000 to \$100,000 [4]. These costs forced significant financial burdens on researchers regarding robots' acquisition expenses.

To solve the cost issue and allow more reproducibility tests for reliability and confidence, augmented reality (AR) can offer a solution. AR adds situated virtual overlays onto our physical world [7], and in our case, AR allows us to situate a wholly virtual robot in its physical task environment. This direction has been partially shown promising in [15]'s work in which participants rated a virtual AR arm mounted on a physical mobile robot the same regarding task performance, socially present, competent, warm, and likable to the more expensive physical ones when using deictic gestures; Indeed, the virtual AR robot arm was rated more anthropomorphic due to the lack of mechanical sound. Although promising, their study has focused on the arm appendage and reference behaviors. It remains unknown if a whole AR virtual robot can retain these perceptions or in common manipulation and navigation tasks, limiting the ecological validity required to use the results confidently in general real-world tasks.

In this work, we designed an experiment to compare the performance of a physical robot and the AR counterpart subjectively to complete an identical task in physical settings. Future results will inform whether we can use an AR virtual robot in a physical environment for some studies instead of a physical robot. Another benefit is enhancing accessibility in HRI experiments, primarily focusing on the cost of HRI experiments with physical robots. We

expect to reduce the economic barrier with AR, which can replace the high-cost physical robot with its AR virtual version.

2 RELATED WORK

Over the past few decades, HRI research has shown differences in perception comparing purely virtual and physical robots. For example, physical robots are better at influencing [9], learning [19], enjoyability [10], and proximity [9].

As AR progresses and integrates more into HRI, exploring the potential impact becomes important. In contrast to virtual agents that wholly exist on computer screens, AR enables virtual objects or agents to be displayed over a user's perception of reality [7]. Multiple research examined how individuals perceive AR agents mixed into the physical world. [25] found that participants perceive AR agents as physically distant when altering the audio volume level according to distance. Furthermore, [20] showed that properly occluding virtual humans increases the co-presence, i.e., the sense of being together, and makes the virtual human's behavior more physically plausible in AR settings.

In HRI, many studies have examined how AR can visualize its behaviors and intent, according to a recent survey by [32]. As mentioned earlier, most relevant to our study is the work done by [16], who mounted an AR virtual robot onto a mobile robot and compared it to physical counterparts for making pointing gestures. Results showed that a more cost-effective virtual arm in AR was on par with the physical counterpart in the experiment. Specifically, the AR virtual arm was rated equally socially present, competent, warm, likable, and more anthropomorphic.

Besides the intersection of AR and HRI, we now discuss the related work done in replicability and reproducibility. In recent years, the need to make HRI studies more reproducible and replicable has attracted more and more attention [8, 12–14, 21, 28]. Notably, [8] proposed developing an online database with a standardized form to assist HRI researchers in documenting and sharing the specifics of their studies. This resource aims to facilitate the replication of HRI research. Moreover, [12] accessed 414 papers of HRI and RO-MAN conferences through 2019–2021, revealing that over 62% papers needed to report recruitment methods and compensation. This research offered recommendations for enhancing reporting study metadata for more reproducible studies in HRI. [14] also highlighted that the training of HRI researchers could encourage the reproducibility of HRI studies. Most recently, [27] open-sourced a generalized language-based experiment software project using the most popular NAO/Pepper platform and shared a concrete workflow, hopefully overcoming replicability barriers to compare the results with the same robot platform studies in HRI research.

Regarding VR to improve applicability and reproducibility, many studies [22, 23, 30, 33] explored replacing physical robots with VR robots in experiments. For example, [22] compared the difference between VR and physical robots interacting with humans and found that humans feel less comfortable interacting with VR robots regarding personal space preferences. [30] showed that VR could successfully replicate the results of interactive experiences in virtual environments with the same complexity as in real environments. However, the interaction between the virtual robot and the real world is a limitation that cannot be ignored.

Thus, it is yet unknown whether we can use AR virtual robots as low-cost substitutes for physical robots in empirical studies. Our work investigates whether the findings from AR robots can transfer to physical robots, thus enhancing reproducibility and accessibility.

3 HYPOTHESIS

To ensure our findings from comparing physical robots and virtual robots in a physical environment generalize to a wide variety of works in HRI, we used the most common measures found in [35]'s work and developed the following hypotheses. Specifically, they [35] analyzed over 1400 papers from the ACM/IEEE HRI and the IEEE RO-MAN conference between 2015 and 2021 to classify subjective and objective measures that could enhance the replicability and usability of HRI research. Our study adopted the research findings to evaluate subjective experience. At submission time, we are still working on hypotheses about objective performance.

Equal subjective experience: We believe the AR virtual robot situated in a physical environment will have equal or greater positive perceptions as a physical robot, measured by usability, trust, and personal preference. This is grounded in the AR robot appendage study by [16].

4 METHOD

To test our hypothesis, we designed a within-subjects experiment.

4.1 Apparatus

4.1.1 Robot Platform. For our results to be generalizable to both manipulation and navigation tasks and compatibility with the widely adopted Robot Operating System (ROS) [26], we plan to use a Fetch mobile manipulator [34]. The robot has a 7-degree-of-freedom arm centrally positioned in front of the torso. Its base circumference measures 60cm. Its torso can adjust height between 1.1m and 1.5m. Its arm extends to a maximum of 114cm, diminishing to 83cm with a downward-pointing gripper. Considering the arm's chest position and the base's radius, effective reachability is limited to approximately 53cm.

Fetch also has a Primesense Carmine RGBD camera, ranging from 0.35m to 1.4m. We will use the robot's ROS package [2] for the virtual robot and render it in Unity while controlling it with MoveIt through ROS#, similar to what [16] did. Thus, we ensure constant variables that the virtual Fetch robot will have exactly the same size, appearance, and movement characteristics as the physical Fetch.

4.1.2 Augmented Reality Head-Mounted Display. We use a Microsoft HoloLens 2 [1]: a commercially available holographic, optical see-through AR display with a field of view of $43^\circ \times 29^\circ$.

4.1.3 Task Environment. As shown in Figure 1, the task environment has multiple tables (stations) with different parts and containers for the Fetch robot to collect and place into corresponding caddy sections. It also has a gear machine to create threads from raw gears, which we do not use as the physical machine is unavailable.

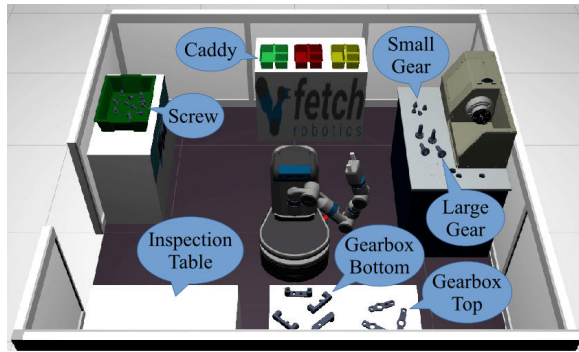


Figure 1: The FetchIt! Mobile Manipulation Challenge environment in simulation. The robot aims to place a specific set of parts into different caddy sections and then carry the caddy to the inspection table for human assembly.

4.2 Task and Implementation

We plan to let Fetch navigate and manipulate (pick-and-place tasks) between three stations (Gear Station, Caddy Station, and Inspection Station). We will use the apparatus introduced in Section 4.1.

The Fetch robot will interact with three types of objects. It will pick, carry, and place the objects between three stations to complete the task. We plan to randomly place six small and four large gears at the gear station, and two caddies will be placed at the caddy station. The Fetch robot will first move to the gear station. After arriving, the robot will identify and distinguish between the gear bottoms and gear tops. The robot will pick up one gear bottom and carry it to the caddy station. Upon arrival, the robot will detect the large compartment in the caddy, where it will place the gear. After this, the robot will transport the caddy to the inspection station.

All objects will be physical in this experiment except for the object to be manipulated because an AR virtual robot can not manipulate physical objects. As there are already CAD models in the Gazebo (a simulation software) simulation environment [6], we will import them to Unity to create the virtual representation of the manipulated object, aligning with the dimensions of the actual physical experimental space.

When the virtual Fetch robot places a gear into the physical caddy, we will apply occlusion techniques to replicate the level of coherence found in the physical world. This technique ensures that participants will not see what is inside the physical caddy, similar to the constraints in the physical world.

4.3 Experiment Design

We will implement three conditions and counterbalance the ordering effect by applying a Latin square to this within-subjects study. We will manipulate the physicality (physical vs. virtual) of the robot and the objects to be manipulated. Thus, we have:

- 1) The physical robot manipulates physical objects.
- 2) The physical robot manipulates virtual objects.
- 3) The virtual robot manipulates virtual objects.

Note that theoretically there is a fourth condition: the virtual robot manipulating physical objects. However, we remove it because it is practically impossible.

4.4 Procedure

We will conduct this study in person, as participants will wear the HoloLens 2 AR headset. Upon arrival, participants will receive an informed consent form. Following consent to participate, they will complete a demographic survey. Then, they will watch videos showing how to wear the HoloLens 2 and calibration. Afterward, they will wear the headset and be allocated to one of three experimental conditions and experience the other two according to the Latin square sequence assigned to them. As this is a within-subjects study, participants will wear the AR headset throughout the study to avoid not wearing it confounding the study results.

Before experiencing the first condition, we will ask questions to confirm participants' understanding of the task and the procedure. Upon experiencing each condition, participants will be asked to complete a questionnaire about all subjective measures. Finally, they will be debriefed and receive payment to compensate for their time. We will pilot the study, and based on the average completion time, we will calculate the compensation accordingly.

4.5 Data Collection and Measures

Partially drawing inspiration from [35]'s work on top HRI measures, we will use three subjective metrics to test our hypotheses. Specifically, we will measure usability using the System Usability Scale (SUS) [11], the most frequently used named survey to measure usability. For assessing trust, because [35] showed that the most commonly used named survey is the Trust in Automation Scale [18], which was only used three times. We will instead use the more recent yet already widely cited Multi-Dimensional Measure of Trust (MDMT) [24, 29]. In short, MDMT captures both the performance and moral aspects of trust. Lastly, we will ask which robot type the participants prefer to measure personal preference.

4.6 Data Analysis

We will analyze our data using the Bayesian analysis framework [31] rather than the Frequentist approach. Bayesian hypothesis testing allows us to quantify evidence for and against competing hypotheses. This means that it can quantify evidence in favor of a lack of an effect (\mathcal{H}_0), particularly fitting the test of our hypothesis of equal subjective experience. (Note that the Frequentist approach cannot support the null hypothesis.) Specifically, the Bayesian approach uses the Bayes Factor (BF), a ratio of evidence between the two competing hypotheses \mathcal{H}_1 and \mathcal{H}_0 . For example, $BF_{10} = 5$ means that the data collected is five times more likely to occur under \mathcal{H}_1 than \mathcal{H}_0 , thus supporting \mathcal{H}_1 . For more details and the benefits of the Bayesian approach, we refer readers to [31].

4.7 Participants

As the Bayesian approach is not grounded in the central limit theorem, it does not require power analysis to ensure the validity of the statistical data analysis, unlike the frequentist approach. Nonetheless, we plan to post posters on bulletin boards and mailing lists to recruit at least 30 individuals from the University of South Florida community, including students, faculty, and nearby residents.

5 INQUIRIES OR DISCUSSION FOR MENTOR

As mentioned earlier, we are currently still investigating objective measures. As an AR virtual robot could not manipulate physical objects, we want to discuss what objective metric we would collect data to measure them, and what hypotheses we may propose. Moreover, I (the first author) want to discuss more about some other common measures that are important for our study results to be generalizable that are not in [35]'s work. Specifically, we believe [35]'s work is limited in that, for trust, MDMT was used two times in the previous conference papers. The Trust in Automation Scale [18], another named survey, was used three times. This shows the limitation by limiting us to the HRI and RO-MAN literature from 2015-2021.

6 CONCLUSION

Motivated by the cost-effective alternative of using AR robots in physical environments for conducting HRI studies, as well as promoting replicability and reproducibility, we designed an experiment to compare the subjective perception of AR virtual and physical robots in physical environments completing identical tasks. To ensure the generalization of our future findings, we focus on measuring usability, trust, and personal preference, which are the most investigated subjective measures. We expect the findings will suggest that AR offers a cost-free, replicable, and accessible option for HRI research, indicating the potential for broader application in studies with AR virtual robots yet physical environments where budget and replicability are key considerations.

REFERENCES

- [1] [n. d.]. About HoloLens 2. <https://learn.microsoft.com/en-us/hololens/hololens2-hardware>. Accessed: 2024-1-14.
- [2] [n. d.]. Fetch & Freight Manual - Fetch & Freight Research Edition Melodic documentation. <https://docs.fetchrobotics.com>. Accessed: 2024-2-16.
- [3] [n. d.]. Hello Robot: Open Source Mobile Manipulator for AI & Robotics. <https://hello-robot.com/>. Accessed: 2024-2-16.
- [4] [n. d.]. HomeRobot: An Open Source Software Stack for Mobile Manipulation Research. <https://drive.google.com/file/d/1-fNMoTJ9pwEVP9RqqV53oZrLmcGdAYS/view>. Accessed: 2024-01-14.
- [5] [n. d.]. ROS Research Arms. <https://www.trossenrobotics.com/robotic-arms/ros-research-arms.aspx>. Accessed: 2024-2-16.
- [6] [n. d.]. ZebraDevs/fetch_gazebo: Gazebo simulator for Fetch. https://github.com/ZebraDevs/fetch_gazebo. Accessed: 2024-2-17.
- [7] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: teleoperators & virtual environments* 6, 4 (1997), 355–385.
- [8] Shelly Bagchi, Patrick Holthaus, Gloria Beraldo, Emmanuel Senft, Daniel Hernandez Garcia, Zhao Han, Suresh Kumar Jayaraman, Alessandra Rossi, Connor Esterwood, Antonio Andriella, et al. 2023. Towards Improved Replicability of Human Studies in Human-Robot Interaction: Recommendations for Formalized Reporting. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 629–633.
- [9] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3 (2011), 41–52.
- [10] Christoph Bartneck. 2003. Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*. 55–60.
- [11] John Brooke. 1996. Sus: a 'quick and dirty' usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.
- [12] Julia R Cordero, Thomas R Groechel, and Maja J Matarić. 2022. A review and recommendations on reporting recruitment and compensation information in HRI research papers. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1627–1633.
- [13] Marlena R Fraune, Iolanda Leite, Nihan Karatas, Aida Amirova, Amélie Legeleux, Anara Sandygulova, Anouk Neerinx, Gaurav Dilip Tikas, Hatice Gunes, Mayumi Mohan, et al. 2022. Lessons learned about designing and conducting studies from hri experts. *Frontiers in Robotics and AI* 8 (2022), 772141.
- [14] Hatice Gunes, Frank Broz, Chris S Crawford, Astrid Rosenthal-von der Pütten, Megan Strait, and Laurel Riek. 2022. Reproducibility in human-robot interaction: furthering the science of HRI. *Current Robotics Reports* 3, 4 (2022), 281–292.
- [15] Zhao Han, Albert Phan, Amia Castro, Fernando Sandoval Garza, and Tom Williams. 2022. Towards an Understanding of Physical vs Virtual Robot Appedage Design. In *International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction*.
- [16] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. 2023. Crossing Reality: Comparing Physical and Virtual Robot Deixis. In *2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.
- [17] Guy Hoffman and Xuan Zhao. 2020. A primer for conducting experiments in human-robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 1 (2020), 1–31.
- [18] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics* 4, 1 (2000), 53–71.
- [19] James Kennedy, Paul Baxter, and Tony Belpaeme. 2014. Children comply with a robot's indirect requests. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 198–199.
- [20] Hanseob Kim, Myungho Lee, Gerard J Kim, and Jae-In Hwang. 2021. The impacts of visual effects on user perception with a virtual human in augmented reality conflict situations. *IEEE Access* 9 (2021), 35300–35312.
- [21] Benedikt Leichtmann, Verena Nitsch, and Martina Mara. 2022. Crisis ahead? Why human-robot interaction user studies may have replicability problems and directions for improvement. *Frontiers in Robotics and AI* 9 (2022), 838116.
- [22] Rui Li, Marc van Almkerk, Sanne van Waveren, Elizabeth Carter, and Iolanda Leite. 2019. Comparing human-robot proxemics between virtual reality and the real world. In *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 431–439.
- [23] Oliver Liu, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2017. Understanding human-robot interaction in virtual reality. In *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 751–757.
- [24] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*. Elsevier, 3–25.
- [25] Mohammad Obaid, Radoslaw Niewiadomski, and Catherine Pelachaud. 2011. Perception of spatial relations and of coexistence with virtual agents. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*. Springer, 363–369.
- [26] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.
- [27] Uthman Tijani, Hong Wang, and Zhao Han. 2024. Towards Reproducible Language-Based HRI Experiments: Open-Sourcing a Generalized Experiment Project. In *2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.
- [28] Daniel Ullman, Salomi Aladia, and Bertram F Malle. 2021. Challenges and opportunities for replication science in hri: A case study in human-robot trust. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*. 110–118.
- [29] 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*. 263–264.
- [30] Valeria Villani, Beatrice Capelli, and Lorenzo Sabatini. 2018. Use of virtual reality for the evaluation of human-robot interaction systems in complex scenarios. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 422–427.
- [31] Eric-Jan Wagenmakers, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Jonathon Love, Ravi Selker, Quentin F Gronau, Martin Šmíra, Sacha Epskamp, et al. 2018. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review* 25 (2018), 35–57.
- [32] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafir. 2023. Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–39.
- [33] Luc Wijnen, Séverin Lemaignan, and Paul Bremner. 2020. Towards using virtual reality for replicating HRI studies. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 514–516.
- [34] Melonee Wise, Michael Ferguson, Derek King, Eric Diehr, and David Dymesich. 2016. Fetch and freight: Standard platforms for service robot applications. In *Workshop on autonomous mobile service robots*. 1–6.
- [35] Megan Zimmerman, Shelly Bagchi, Jeremy Marvel, and Vinh Nguyen. 2022. An analysis of metrics and methods in research from human-robot interaction conferences, 2015–2021. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 644–648.