An interdisciplinary approach to intelligent disobedience: A Nuanced Exploration of User Experience in Human-Induced Interaction Failures during Teleoperation

Kavyaa Somasundaram
kavyaa.somasundaram@oru.se
Örebro University
Örebro, Sweden

Katherine Harrison
katherine.harrison@liu.se
Linköping University
Linköping, Sweden

Andrey Kiselev
andrey.kiselev@oru.se
Örebro university
Sweden

Amy Loutfi
amy.loutfi@oru.se
Örebro university
Sweden

ABSTRACT

Failures can occur in any interaction between humans and robots, and the extent of the failure depends on the domain of application and the context in which the failure occurs. Typically, interaction failures are attributed to errors on the robot's side. This paper examines the concept of Intelligent Disobedience (ID) and how it can be leveraged for managing interaction failures caused by humans. Our work-in-progress focuses on the practical implementation of ID in robot teleoperation and highlights the benefits of interdisciplinary collaboration for this project.

CCS CONCEPTS

• **Computer systems organization → Robotics**; • **Human-centered computing → Interaction design theory, concepts and paradigms**.

KEYWORDS

intelligent disobedience, failures in HRI, interaction failures

ACM Reference Format:
Kavyaa Somasundaram, Katherine Harrison, Andrey Kiselev, and Amy Loutfi. 2023. An interdisciplinary approach to intelligent disobedience: A Nuanced Exploration of User Experience in Human-Induced Interaction Failures during Teleoperation. In *Proceedings of Interdisciplinary Approaches in Human-Agent Interaction workshop (Inter HAI WS '23).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The field of Human-Robot Interaction (HRI) is dedicated to designing robots that can engage with humans instinctively and naturally [1]. The emergence of teleoperated robots, with some level of autonomy, is becoming prevalent in areas where there is a vital need for humans and robots to work side by side to achieve a common goal. Trust and safety are the central pillars of establishing this collaborative workspace [8]. However, disruptions in this partnership are commonly looked at in terms of inherent robot design issues [6], or miscommunication arising from linguistic or cultural nuances [4]. In this paper, we discuss the less explored cause of disruptions to collaborative human-robot workflow, that is human-induced errors/failures in the teleoperation domain. Within the varied definitions of failures in HRI literature, we follow the definition given by Brooks, " a degraded state of the ability which causes the behavior or service being performed by the system to deviate from the ideal, normal or correct functionality" [2]. Failures are further classified into different categories, where human-induced interaction failures are subdivided into mistakes(conscious decisions), lapses (attention/memory lapses), slips (unconscious decisions), and deliberate violations (voluntary intention to disrupt the task) [6]. Given the numerous possibilities for human-induced interaction failures, there is a pressing demand to devise and enhance systems that can proficiently manage such failures, ensuring smooth human-robot synergy. In this paper, we discuss the development of an interdisciplinary methodology designed to explore the impact of such human-induced errors. We provide an overview of the experiments and interdisciplinary activities that have been planned for our ongoing and upcoming research studies.

2 INTELLIGENT DISOBEDIENCE: AN APPROACH FOR HUMAN-INDUCED INTERACTION FAILURE MANAGEMENT

The concept of Intelligent Disobedience (ID) is adapted from the training methods used for guide dogs. Guide dogs serve as a companion to individuals with visual impairments, assisting them in their daily activities, particularly in navigation. These dogs are trained to disregard their handler's orders if deemed unsafe to execute. E.g., refusing to proceed forward when a cyclist is approaching in the same direction. In teleoperation, this principle is applied by programming the robot/system to anticipate and reject potentially harmful commands [3]. In summary, the act of ID is doing right when what you are instructed to do is wrong or dangerous.

Inspired from [7], we proposed an architecture to manage human-induced interaction failures during a teleoperation scenario ([9], [10]) by categorizing it into three stages:

Step 1: Identification of failure triggers Detecting conflicts between the global objectives (robot knowledge based on the priorities of the task) and the local objectives (commands issued by the operator). Then, tagging that activity as a failure trigger.

Step 2: Communication of feedback After the system identifies the failure triggers, it is crucial to return that information to the operator. This step aims to increase transparency and trust in the system.

Step 3: Mitigation of failures The last step is to mitigate this failure trigger by choosing the appropriate tools based on the robot's capability. Mitigation strategies range from utilizing the control mechanisms (obstacle avoidance, depth holding, etc.,) to complete takeover of control from the user.

In the previous work,[9] the authors also suggested that ID represented an area in which the need for a more nuanced understanding of human-robot relations and an accompanying ethical framework to guide such research was necessary. In what follows we outline an ongoing attempt to develop such an understanding through an interdisciplinary collaboration between robotics and gender studies.

2.1 Experiment Outline

To evaluate the feasibility and outcomes of implementing ID on a robot, we are conducting a proof of concept simulation experiment using a wizard-of-oz approach. The experiment involves testing the usability of ID in a teleoperation scenario, where users are required to navigate a robot through unstructured environments by following markers and reaching the goal point. The experiment consists of three settings such as, without wizard interruptions, with wizard interruptions, and finally with wizard interruptions and feedback. Post experiment the users are asked to fill in standard questionaires to quantify trust and workload across different runs.

The main goal here is to test and measure the difference in trust and workload quantitatively across the above-mentioned experiment setting and to gain insights on the usefulness and after-effects of implementing ID onto a robot.

3 NUANCING ID

Intelligent disobedience has a clear use in potentially dangerous situations in which a robot's perception of the environment might be superior to a human's. However, it also opens up an ethically uncertain area in terms of determining when and how a robot should disobey and demands sensitivity to the expertise and vulnerabilities of all those involved [5]. Furthermore, attitudes towards (dis)obedience are socioculturally specific and may indicate differential power relations rooted in age, gender, ability, or socioeconomic status [11]. Gaining a deeper understanding of how human users perceive and respond to moments of ID is thus essential to designing a better interaction.

With this in mind, this paper advocates an early attempts to develop an interdisciplinary methodology for the ID experiment (outlined in the previous section). Quantitative data collected through standard questionnaires as well as operator performance data from the interface will be collected before, during, and after the experiment. As an attempt to promote interdisciplinary works, qualitative data analysis and data collection are planned. Qualitative data is to be collected during the experiment through observations, and after the experiment, through semi-structured interviews. Qualitative data collection will focus on sensitizing the experiment to differences in experience based on previous experience with robots as well as sociocultural factors. This approach will (1) help us gain a deeper understanding of the impact of ID on trust, (2) verify if the results observed through quantitative data are consistent with the observations made during the interviews, (3) propose new avenues of development and (4) contribute to wider long-term dialogue around qualitative-quantitative collaborations for enriched data collection in HRI.

4 CONCLUSION

The use of ID technology in teleoperation is a recent development that holds immense potential for creating robots that are safe, and proactive. However, the practical implications of implementing ID require a carefully designed system that takes into account the ethical framework of the application domain. To achieve effective long-term interactions, interdisciplinary studies, like the one outlined in this paper, are crucial. Evaluating trust both quantitatively and qualitatively provides key insights into the human's interpretations and expectations towards a system equipped with ID. Questions such as who takes responsibility for incorrect decisions made by the system, and how social factors impact adoption of technologies, are difficult to answer but must be considered. We approach these issues by actively participating in interdisciplinary studies and raising awareness of the implications and risks involved. We seek and encourage opportunities to open up discussions on the challenges involved in developing and implementing ID by inviting people from different specializations.

5 ACKNOWLEDGMENTS

This work has been supported by the Industrial Graduate School Collaborative AI and Robotics funded by the Swedish Knowledge Foundation DNR:20190128. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation.

REFERENCES

[1] C. Breazeal. 2004. Social interactions in HRI: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, 2 (2004), 181–186. https://doi.org/10.1109/TSMCC.2004.826268

[2] Daniel J Brooks. 2017. HUMAN CENTRIC APPROACH TO AUTONOMOUS ROBOT FAILURES. *Ph.D. dissertation, Department of Computer Science, University. Lowell, MA* (2017), 24.

[3] Ira Chaleff. 2015. *Intelligent disobedience: doing right when what you're told to do is wrong* (first edition ed.). Berrett-Koehler Publishers, Inc, Oakland, CA.

[4] Petra Gieselmann and Mari Ostendorf. 2007. Problem-Sensitive Response Generation in Human-Robot Dialogs. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics, Antwerp, Belgium, 219–222. https://aclanthology.org/2007.sigdial-1.38

[5] Katherine Harrison and Ericka Johnson. 2023. Affective Corners as a Problematic for Design Interactions. *J. Hum.-Robot Interact.* 12, 4, Article 41 (jun 2023), 9 pages. https://doi.org/10.1145/3596452

[6] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in Psychology* 9, JUN (2018). https://doi.org/10.3389/fpsyg.2018.00861

[7] Reuth Mirsky and Peter Stone. 2021. The seeing-eye robot grand challenge: Rethinking automated care. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* 1, Aamas (2021), 28–33. ISBN: 9781713832621.

[8] Aaron Pereira, Mareike Baumann, Jonas Gerstner, and Matthias Althoff. 2022. Improving Efficiency of Human-Robot Coexistence While Guaranteeing Safety: Theory and User Study. *IEEE Transactions on Automation Science and Engineering* (2022), 1–14. https://doi.org/10.1109/TASE.2022.3201970 Conference Name: IEEE Transactions on Automation Science and Engineering.

[9] Kavyaa Somasundaram, Katherine Harrison, Amy Loutfi, and Andrey Kiselev. 2023. Nuancing the human-robot relation: Intelligent disobedience and human failures in robot teleoperation. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, ImpRR23 workshop* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA. https://sites.google.com/view/hri-failure-ws/program

[10] Kavyaa Somasundaram, Andrey Kiselev, and Amy Loutfi. 2023. Intelligent Disobedience: A Novel Approach for Preventing Human Induced Interaction

Failures in Robot Teleoperation. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 142–145. https://doi.org/10.1145/3568294.3580060

[11] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/3568162.3576973