The Unspoken Rules of Engagement: Understanding the Impact of Social Norm Violations in Human-Robot Interactions

by

Steven Lawrence

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2023

© Steven Lawrence 2023

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public

Statement of Contributions

Steven Lawrence was the sole author for Chapters 1, 3, and 7 which were written under the supervision of Professors Kerstin Dautenhahn, Chrystopher Nehaniv, and Jesse Hoey. This thesis includes first-authored peer-reviewed material published in conference proceedings by the Institute of Electrical and Electronics Engineers (IEEE), as well as other work submitted for publication.

Research presented in Chapters 2, 4, 5, and 6:

In Chapter 2, Steven Lawrence, and Dr. Melanie Jouaiti conducted a systemic review of social norms in HRI. Steven Lawrence and Dr. Jouaiti jointly performed all screening conditions and data extraction required for the review. Steven Lawrence comprised the review from the extracted data.

In Chapters4, 5, and 6, Steven Lawrence, Negin Azizi, Kevin Fan, and Dr. Jouaiti conducted an HRI user study. Steven Lawrence and Dr. Jouaiti designed the experiment. Steven Lawrence implemented system requirements to conduct the experiment. Steven Lawrence, Negin Azizi, and Kevin Fan collaboratively conducted the in-person study.

All authors reviewed and edited manuscripts.

Citations:

Steven Lawrence, Negin Azizi, Kevin Fan, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. The impact of social norm violations on participants' perception of and trust in a robot during a competitive game scenario. Accepted for publication in Proc. IEEE RO-MAN. IEEE, 2023. In press

Steven Lawrence, Negin Azizi, Kevin Fan, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. Exploring the nuances: Trust, discomfort, and physiological changes in response to robotic norm violations. Manuscript submitted for publication, 2023.

Steven Lawrence, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. The Role of Social Norms in Human-Robot Interaction: A Systematic Review. Manuscript submitted for publication, 2023.

Abstract

As robots increasingly permeate diverse domains such as healthcare, education, service industries, and domestic environments, the significance of understanding and navigating human-robot interactions intensifies. This thesis explores the role of social norms in human-robot interaction (HRI), specifically investigating the effects of norm-violating behaviours by robots on trust, discomfort, competence, enjoyment, and physiological state changes in human participants. To begin this research, we conducted a systematic review to gain a comprehensive understanding of existing literature on social norms in HRI and identify any potential research gaps we could fill. Our analysis revealed that social norms play a significant role in shaping user perceptions and experiences in HRI, with robots that adhere to social norms generally being perceived as more acceptable, trustworthy, and effective. Moreover, future research should involve direct interactions with robots rather than relying on online studies and simulations, and extend beyond the reliance solely on self-reported questionnaires by incorporating diverse assessment methods (*e.g.,* behavioural measures and physiological data) to capture the intricacies of human-robot interactions.

Our experiment used a 2x1 between-participant experimental design where participants were randomly assigned to one of two groups (*i.e.,* experimental or control). The participants in the experimental group were exposed to social norm violations carried out by a robot during a competitive scavenger hunt game. In contrast, the participants in the control group did not experience any social norm violations by the robot. Throughout the experiment, we collected video footage of the participants for behavioural observations, and physiological data in the form of Galvanic Skin Response (GSR) and Photoplethysmography (PPG) to detect event-driven changes in Skin Conductance Response (SCR) and Heart Rate Variability (HRV). Additionally, we collected responses to self-reported questionnaires and open-ended questions targeting perceptions of trust, discomfort, competence, and overall enjoyment. By integrating diverse data streams we provide a comprehensive account of HRI dynamics in the context of social norm violations. Results reveal significant shifts in human perception and attitudes toward robots when social norms are violated. Notably, we found a decrease in participants' trust and overall enjoyment, an increase in discomfort, and physiological state changes that complement the findings on perceptual changes and further underscore the impact of these norm violations. These findings highlight the importance of adherence to social norms in designing and programming robots to integrate into human-centric environments successfully. This study offers insights into advancing the field of HRI and the design of socially compliant robots. It spotlights the significant influence of norm violations on human perception and physiological states, paving the way for improved understanding and encouraging more positively perceived interactions with robots.

Acknowledgements

This journey of research has been guided and enriched by the immense knowledge, steadfast support, and insightful advice of Professor Kerstin Dautenhahn, Professor Chrystopher L. Nehaniv, and Professor Jesse Hoey. Their influence has been instrumental in shaping the progress and development of this work.

In the same vein, the invaluable assistance, guidance, and rigorous scrutiny from Taif Anjum and Alexander Aroyo have significantly amplified the quality and rigour of this research. Their unwavering commitment and insightful suggestions have been a cornerstone in the construction of this study, and for this, I express my profound gratitude.

Furthermore, I would like to acknowledge the invaluable support and encouragement from my family and friends, who have been a constant source of motivation and inspiration throughout this journey.

Dedication

To Hannah and Archer, my source of love and inspiration. Your unwavering support and presence have made this journey possible. This achievement is dedicated to both of you, with heartfelt gratitude for your endless encouragement and belief in me.

Table of Contents

List of Figures ix

List of Tables xi

1 Introduction 1

1.1 Motivation . . . 1

1.2 Research Objectives and Challenges . . . 2

1.3 Contributions and Thesis Outline . . . 3

2 Systematic Literature Review 5

2.1 Purpose . . . 5

2.2 Methodology . . . 6

2.2.1 Terminology and Scope of the Review . . . 6

2.2.2 Data Search and Categorization . . . 9

2.3 Results . . . 14

2.4 Discussion . . . 20

2.4.1 Conformity of Norms and Real-world Applications . . . 20

2.4.2 Violation of Norms and Potential Concerns . . . 21

2.4.3 Challenges and Broader Applicability . . . 22

2.5 Research Gaps and Future Direction . . . 23

3 Related Background and Prior Works **25**

3.1 Definitions and The Role of Social Norms in HRI 25

3.2 Social Norm Adherence and Violation in HRI 27

4 Methodology: A Mixed-Method Approach **30**

4.1 Study Design 30

4.2 System Implementation 33

4.3 Procedure 34

5 Results of User Study **42**

5.1 Participant Perception and Severity Assessment of Norm Violations 42

5.2 Perceptual, behavioural, and Physiological Responses to Norm Violations . 44

5.2.1 Trust, Discomfort, Competence, and Enjoyment 44

5.2.2 behavioural Responses 48

5.2.3 Physiological Responses 49

6 Discussion and Implications **53**

6.1 Preliminary Analysis and Evaluation of Participant Responses 53

6.1.1 Questionnaire Analysis: Assessing Trust, Discomfort, Competence, and Overall Enjoyment 54

6.1.2 Examining Behavioural Responses 56

6.1.3 Examining Physiological State Changes 57

6.2 Interpretation of Findings and Future Directions 59

7 Limitations and Conclusion **61**

References **64**

APPENDICES **80**

A Ethics Approval **81**

List of Figures

2.1 Diagram illustrating the data collection process following the PRISMA guidelines. 10

2.2 The number of articles identified in the systematic review per year. Search includes articles up until February 14, 2023. 12

2.3 The number of articles per norm, with proportionality shown for both the violation and conformity categories. 13

4.1 Two figures depicting the lab environment, and the scavenger hunt game setup including the positions of objects, waiting locations for both the robot and participant, and off-limits areas within the laboratory layout. 31

4.2 A figure showcasing the experimental setup displaying various household and office items and multiple objects, such as boxes, to further conceal the scavenger hunt items. Additionally, it shows different scenarios involving the Fetch robot in the task. 35

4.3 This figure presents screenshots of the user interface that participants interacted with during the game. 36

4.4 Sensor placement of the participant's non-dominant hand from back and front. 39

5.1 The figure presents the percentage of norm violations perceived by participants, the severity ratings, the impact on participants' perceptions of the robot (*i.e.,* cognitive appraisal), and the degree to which it discomforted or agitated them (*i.e., disconcertment*). 44

5.2 The outcomes of both the RIS and ENJOY questionnaires, displaying the relative scales for each on the y-axis. 46

5.3 The mean scores for the RoSAS questionnaire's subscales depict the participants' relative discmfort, warmth, and competence towards the robot. . . 47

5.4 Differences in behavioural observations captured during the analysis of the experimental and control groups. The three categories of behavioural responses—verbal, physical, and emotional—are shown. 48

5.5 Comparison of Heart Rate Variability (HRV) in the two conditions, measured by RMSSD (Root Mean Square of Successive Differences). 51

5.6 Comparison of Galvanic Skin Response (GSR) using Skin Conductance Response (SCR) peak Frequency. 52

List of Tables

2.1 Search Strategy for Literature Review 9

5.1 Pre-questionnaire analyses 43

5.2 Results from post-questionnaires measuring trust, social attributes, user experience, and enjoyment towards the robot in the experimental and control groups. 45

Chapter 1

Introduction

The growing pervasiveness of robots in industries like healthcare, education, service industries, and domestic environments has ignited interest and excitement about the potential of human-robot interactions in the years to come. As robots become more deeply integrated into human-centric environments, exploring and understanding human perceptions and attitudes toward these technological companions is crucial. Social norms are central to these interactions and their subsequent acceptability, which serve as behavioural standards within a group or society [101, 62]. Social norms are instrumental in shaping human interactions and human-robot interactions [88, 15].

However, quantifying the effects of norm-violating behaviours in Human-Robot Interaction (HRI) is challenging and requires careful experimental control. Accurately measuring the impact of norm violations necessitates the integration of diverse data sources to capture the intricacies of HRI [15, 30, 42]. This complexity is further compounded by the growing presence of robots in social environments, emphasizing the importance of developing socially acceptable robots that promote positive human-robot interactions [42, 106].

1.1 Motivation

With robots being integrated into various human-centred environments at an increasing rate, it is crucial to investigate the factors that influence how humans perceive and interact with them. While significant progress has been made in the field of HRI, there is still much to explore regarding the effects of robots' norm-violating behaviours. Robots are prone to making unintentional errors due to challenges in perception and decision-making. These

errors can manifest as violations of social norms, such as bumping into objects or producing unnecessary noise. The consequences of these norm violations on human perceptions and attitudes toward robots require further attention.

Current literature is skewed towards two types of social norm violations, navigation, and communication, leaving other areas under-represented [74]. Our motivation is to address this gap by focusing on the impact of more general social conduct norm violations, *e.g.,* speaking loudly in a quiet or formal setting.

Norm violations in human-robot interactions may affect multiple factors, including trust, discomfort, perceived competence, overall enjoyment, and a person's physiological state. Trust is vital for successful robot implementation, and norm violations have been linked to decreased trust [54]. Further understanding the impact of these violations on trust in HRI is crucial. Similarly, norm violations can lead to discomfort, affecting robot acceptance [42]. Perceived competence and overall enjoyment of the interaction are other critical dimensions that norm-violating behaviours may impact, which are also essential for the long-term acceptance of robots.

To advance the field of HRI and facilitate the successful integration of robots into society, it is necessary to gain a nuanced understanding of the impacts of norm violations on these key aspects. This investigation aims to comprehensively understand the relationship between norm-violating behaviours and human perceptions, ultimately informing the development of socially acceptable robots.

1.2 Research Objectives and Challenges

Building on the existing body of work, we aim to deepen the understanding of the influence of social norms on HRI and to provide further insight into the design and programming of socially acceptable robots. This work aims to examine the impact of norm-violating behaviour on human perception and attitudes toward robots, and a human's physiological state. As such, we formulated the following research questions and hypotheses:

RQ1: How do social norm violations affect trust in robots?

RQ2: How are discomfort, and perceived competence affected by robots performing social norm violations?

RQ3: How is overall enjoyment affected when robots perform social norm violations?

RQ4: Are there physiological state changes (*i.e.,* agitation and stress) when a robot violates social norms that can be detected through the use of physiological sensors (*i.e.,* GSR, PPG)?

Based on the existing literature, we hypothesized that a robot exhibiting social norm violations will **H1:** lower the level of trust in participants [44], **H2a:** increase the level of discomfort in participants [55], **H2b:** negatively affect the perceived competence of the robot [2], **H3:** reduce overall enjoyment [145], and **H4:** elicit physiological state changes [110, 107].

1.3 Contributions and Thesis Outline

In our study we programmed the mobile manipulator robot, Fetch [141], to engage in behaviours that intentionally violated social norms during a competitive scavenger hunt game. This practical and engaging context provided a platform for studying norm violations. The study identified four instances of norm violations, categorized as two *injunctive* and two *descriptive* norm violations (cf. definitions of those below, 3.1). The robot's injunctive norm violations consisted of stealing the participant's object instead of its own and intentionally damaging boxes labelled "fragile" to access an object. The descriptive norm violations included making a disruptive noise in a shared lab environment (participants had previously been briefed on this) and using part of the participant's turn to complete its own turn, with further details provided in Section 4.1. Our approach to studying norm violations in HRI aims to enhance our understanding of how social norms affect users. By gaining insights into these impacts, we hope to better inform the design and programming of social robots sharing spaces with humans, leading to more successful and positive human-robot interactions.

The following are the summarized contributions of this thesis:

1. Utilization of a Comprehensive, Mixed-Method Approach: We combined questionnaires, behavioural analysis, and physiological data to comprehensively understand the impacts of norm violations.

2. Identification of Norm Violations: Our study used four distinct norm violations, two injunctive and two descriptive, offering a more granular understanding of norm violations in human-robot interactions.

3. Impact on User Perceptions: We provide insights into how norm violations influence human perceptions and attitudes toward robots.

4. Systematic Review: Our literature review interprets the current state of the field, and highlights gaps in understanding the impact of norm-violating behaviours on human perceptions and attitudes in HRI.

5. Socially Aware Robot Design: Our findings inform the design and programming of socially aware robots, contributing to more effective human-robot cohabitation.

The structure of this thesis includes Chapter 2, which presents a comprehensive literature review of the current state of research on social norms in HRI. It identifies key themes, and discusses potential research gaps and future directions.

Chapter 3 presents background and literature specifically related to the user study conducted in this thesis. This chapter draws upon the findings of the systematic review, which was conducted to identify and compile the most relevant studies in the field.

Next, Chapter 4 describes the mixed-method approach utilized in this thesis to capture the nuanced impact of norm violations on participants. Furthermore, it outlines the experimental procedure, highlighting the steps taken to ensure the validity and reliability of the study.

Chapter 5 presents the study's findings, covering the results obtained from the questionnaires, behavioural analysis, and physiological data. The chapter analyzes the impact of norm violations on trust, discomfort, perceived competence, overall enjoyment, and physiological state. It provides insights into the consequences of norm violations in HRI.

Moving on to Chapter 6, it covers the discussion of the findings presented in the previous chapter. The chapter explores the implications of the findings, addresses any discrepancies or unexpected outcomes, and provides interpretations and explanations for any observed patterns.

Lastly, Chapter 7 concludes the thesis by summarizing the key findings, highlighting the contributions made to the field, and discussing the limitations of the study. This chapter also discusses the implications of the research findings, their potential impact on the development of socially acceptable robots, and the broader implications for the field of HRI. Finally, it concludes with suggestions for future research to further advance our understanding of norm violations in HRI.

Chapter 2

Systematic Literature Review

2.1 Purpose

The increasing integration of robots into various aspects of our lives, including homes, workplaces, schools, and care for older adults, requires the development of robots that exhibit domain-specific social attributes and behaviours. Despite the growing body of literature on social norms in Human-Robot Interaction (HRI), a comprehensive understanding of their role in robot design remains elusive. This systematic review addresses this gap by examining the existing literature on social norms in HRI and their implications for robot design.

Norms, defined as cultural phenomena that prescribe and proscribe behaviour in specific circumstances, regulate social behaviour [56]. The state enforces some norms as laws, while others are informally enforced by group peers, known as social norms. Adherence to these norms can impact trust [137], acceptance [86], and willingness to help [37], making them crucial factors to consider when designing robots and their behaviours. Therefore, understanding and formalizing human social norms is essential for designing robots capable of norm-conforming behaviour [90]. Developing cognitive and computational models for human representation, activation, and learning of norms is necessary for designing norm-competent social robots [90]. Additionally, artificial agents must meet specific demands, such as representing multiple norm types, context-sensitive activation, and learning and updating norms [89].

Earlier studies have demonstrated that social norms have a considerable impact on human perceptions and attitudes toward robots [88, 15] and that the social attributes of

robots have been shown to influence many aspects of the interaction, such as quality of the interaction [10], interpersonal learning [16], making behaviour interpretable/legible [79], and safety performance [60]. Conversely, a robot lacking social attributes may result in user resistance and minimal engagement [67]. Anthropomorphic design has been used to increase the social attributes of robots, and research has shown that it can improve acceptance and interaction quality between humans and robots [73, 98, 38]. However, some studies have shown that anthropomorphism raises expectations of capabilities and the assumption of norm adherence [124, 147].

Our research question is: "What role do social norms play in HRI, and how can they be incorporated into robot design to improve user experience and acceptance?" To address this question, we systematically reviewed the literature following the PRISMA guidelines [95, 103], employing a rigorous search strategy, inclusion/exclusion criteria, and data extraction process. Our findings provide a comprehensive understanding of the role of social norms in HRI and their implications for robot design, as well as identify research gaps and suggest future directions for the field.

The review revealed that incorporating social norms into robot design poses several challenges, such as accounting for cultural variations, context-specific norms, and evolving norms over time. For example, a robot designed for a specific cultural context may struggle to adapt to another culture's norms, or it may need to update its behaviour as societal norms evolve. Addressing these challenges can improve user experiences, promote broader acceptance of robots, and foster more successful integration of robots in various domains [42, 106]. Studies on social robot acceptance suggest that sociability, adaptability, enjoyment, companionship, and perceived behavioural control are crucial factors to consider when evaluating or creating social robots [32].

As the field of HRI continues to expand, this systematic review aims to provide a comprehensive understanding of the role of social norms in HRI and explore the methodologies, challenges, and potential solutions in developing socially aware robots. By examining the literature, we hope to offer valuable insights and a foundation for future research in the field of HRI.

2.2 Methodology

2.2.1 Terminology and Scope of the Review

For this review, we define a social robot as a type of robot that communicates and harmonizes its behaviour with humans through verbal, nonverbal, or affective modalities,

possessing the social skills to interact with people in human-centric terms [49, 28, 17]. The term *social norms*, or *norms* for shorthand, will be used in this review as an encompassing term to refer to various types of norms, including injunctive norms, descriptive norms, folkways, and mores [101, 25, 40]. It will also include related terms in the literature such as etiquette and convention.

Although defining social norms is challenging, as the definition may change based on the authors' underlying theoretical perspectives, we will use the following definition in this work: *Social norms* are the unwritten, culturally-rooted rules that prescribe and proscribe behaviour in specific circumstances, regulating social behaviour by shaping shared understandings about actions that are obligatory, permitted, or forbidden [56, 102]. They serve as an essential underlying foundation for any social group, particularly human communities, with violations potentially leading to negative reactions from individuals [62]. These negative reactions can take various forms, such as social ostracism, ridicule, or other forms of social pressure. These reactions encourage conformity to norms and discourage their violation, thereby maintaining the established social order and promoting group cohesion. Cultural norms, which similarly regulate behaviour, are specifically examined within a particular cultural group [99].

Furthermore, while considering the search terms for the review, we considered other concepts such as normative behaviour, and subjective norms. These concepts are related but distinct, and it is important to distinguish between them and not use them interchangeably. Normative behaviour refers to actions and behaviours that are expected or considered appropriate within a given social group or culture [1]. Whereas subjective norms refer to an individual's perceptions of what their important others think they should do [1]. Thus, individuals navigate a complex interplay of social, subjective, and cultural norms to determine the normative behaviours suitable within their specific social context.

Inclusion/Exclusion Criteria

We reviewed articles that integrate social norms into robots to explore the effects of conformity and violations of social norms in robots used in human-centric environments. We used several inclusion and exclusion criteria, which are outlined below:

Inclusion criteria: Articles that explore the use of social norms in robots, including robots' ability to learn and apply social norms, cultural and inclusive awareness related to norms, acceptance of robots concerning social norms, social awareness in robots, human-robot cooperation related to social norms, and human-robot interaction preferences related

to robot etiquette and conventions. We adopted a broad definition of a "robot" for this systematic review, which includes:

1. Physical Robots: This category encompasses tangible entities capable of interacting with human users and performing tasks in a physical space, except for autonomous vehicles. The studies within this category involve direct, in-person interactions between participants and robots (*e.g.,* path-crossing, communicating, playing a game).

2. Simulated Robots: These are virtual representations of robots within a simulated environment. Inclusion required the simulated robot to interact with human-controlled agents within the simulation, creating an environment for HRI to occur. Typically, participants were tasked with controlling simulated agents to engage in activities such as playing a game with the robot or completing navigation tasks.

3. Robot Videos: Studies that used video recordings of robots for their experiments were also included. As defined above, these could be videos of any robot interacting with humans or performing tasks. Studies in this category required participants to watch one or more videos and subsequently provide feedback to rate certain aspects.

4. Robot Imagery: Finally, we also included studies that used static images of robots to understand human perceptions, attitudes, or behaviours in response to these images. These studies typically presented participants with a static image of a robot accompanied by a text-based scenario, and participants were then prompted to provide feedback.

This inclusive definition allowed us to capture a broad spectrum of HRI research, providing a comprehensive overview of the role of social norms in this field. All articles meeting these criteria and identified through our search and data collection approaches were included, except those excluded based on the exclusion criteria outlined below.

Exclusion criteria:

- Articles proposing socially aware models without a user study.
- Articles focusing on autonomous vehicles.
- Articles without a robot.
- Articles with a primary focus on morality, or moral judgment, without a clear connection or contribution to understanding social norms.

All articles retrieved through our search queries and data collection process (detailed in the next section) were considered for inclusion unless they met our exclusion criteria.

2.2.2 Data Search and Categorization

Table 2.1: Search Strategy for Literature Review

Search Terms	Source Publications
(social* OR societ* OR cultur* OR subjective) AND (norm OR normative OR etiquette OR robotiquette OR convention) AND (robot)	ACM Digital Library IEEE digital library Scopus Springer Link PsycINFO

Our search found an initial 3227 articles from the following databases: ACM Digital Library, IEEE digital library, Scopus, Springer Link, and PsycINFO. The keywords and phrases used in the search can be viewed in Table 2.1, with the following index terms used where applicable "social norms", and "social robotics". Data was gathered in February 2023 and no exclusion was applied according to the date of publication, an illustration of the number of articles by year can be seen in Fig. 2.2.

In total, 160 in the ACM Digital Library, 790 in Scopus, 314 in the IEEE digital library, 1874 on Springer Link, and 89 on PsycINFO (a total of 3227) papers were identified. The software tool Covidence [134] was then used as a screening platform to detect duplicate articles and perform title/abstract screening, full-text screening, and data extraction. After removing 541 duplicates, 2686 publications remained.

The remaining 2686 articles were checked manually by two reviewers with a Cohen's Kappa [94] inter-rater reliability score of 0.62 (substantial agreement). After performing title and abstract screening, 2482 articles were removed according to our inclusion/exclusion criteria, leaving 204 articles to move forward to the full-text review. The remaining 204 articles were full-text screened, 118 were removed for having no user study, 15 were removed for being irrelevant to the review topic, and 7 were removed for not having a robot, giving a subtotal of 64 articles to be included in the review. An additional 5 articles were sought for retrieval through snowballing via field experts giving a total of 69 articles that successfully met the inclusion/exclusion criteria and were published between 2006-2023. We followed the recommended procedure for systematic reviews, as encapsulated in the PRISMA guidelines [103]. Figure 2.1 provides an overview of the systematic data collection process.

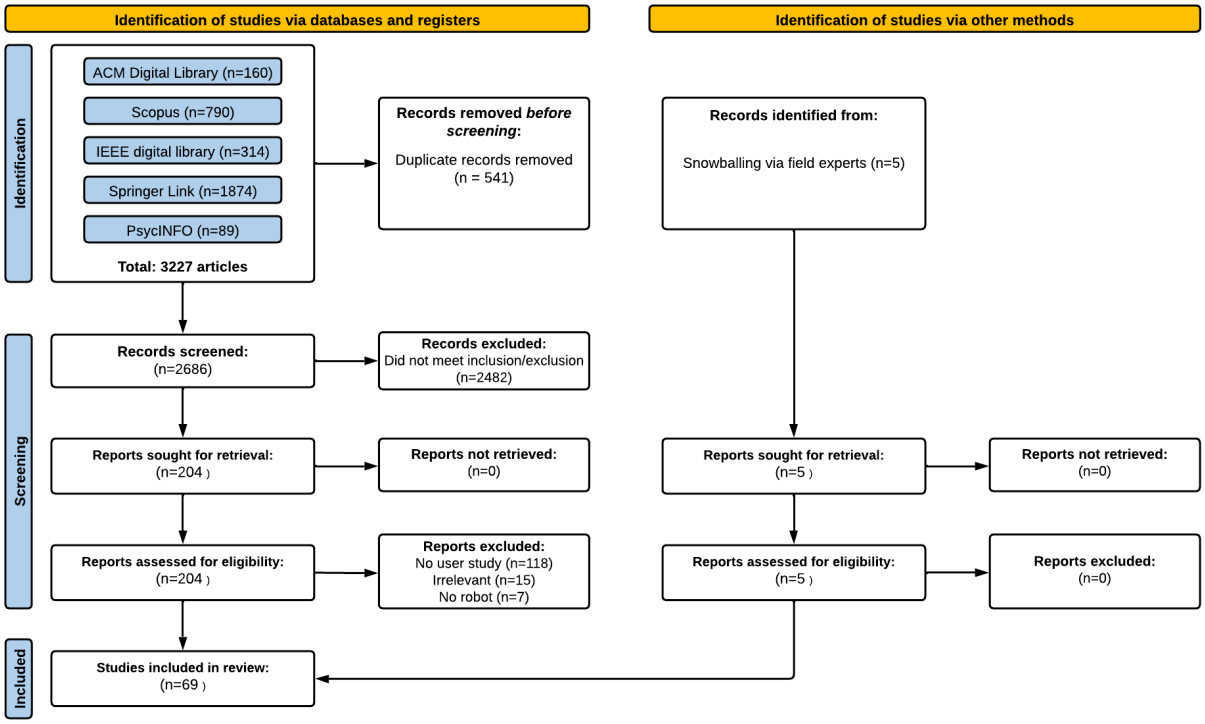


Figure 2.1: Diagram illustrating the data collection process following the PRISMA guidelines. [103]

Categorization

Upon reviewing the selected papers, two main categories emerged, *Norm conformity* (NC) and *Norm violation* (NV):

1. **Norm conformity:** This category includes studies that examine how social norm conformity is incorporated into human-robot interactions. This can include examining how robots conform to social norms, both individual and group norms, as well as how humans respond to robots conforming to these norms. Some studies may also explore how robots can be designed to effectively adhere to social norms, such as turn-taking or polite language, to improve human-robot communication and cooperation. Overall, the norm conformity category focuses on understanding the role of social norms in shaping human-robot interactions and how robots can be designed to conform to these norms.

2. **Norm violation:** This category includes studies investigating the effects of robots violating social norms in human-robot interactions. These studies examine how humans react to robots that violate social norms, both unintentionally and intentionally. The violation of social norms can be in terms of behavioural, verbal, or nonverbal cues, and studies examine the impact of the violation on the perception of the robot, the interaction between humans and robots, and the outcomes of the interaction. The studies in this category also explore the potential factors that can moderate the effects of norm violation, such as cultural differences, individual differences, and contextual factors.

For each paper the following data was collected, (1) the year of the study, (2) the specific robot employed, (3) the task in which the participant(s) were involved, (4) the context of the task that took place, (5) the demographics of the participants involved in the study, (6) the environment in which the study took place, (7) the type of norms the study examined, and finally, (8) the key findings or outcomes derived from the study. Further details for each aspect are provided below.

Tasks and Contexts: These categories encompass the tasks and contextual factors of the experiment that participants were involved in. The task refers to the specific activity or goal participants were asked to perform, while the context refers to the setting where the task occurred. This includes further details of the task, the social situation, and any other relevant contextual factors that may impact the human-robot interaction. It is important to mention that a single study could involve multiple tasks, each with a unique context.

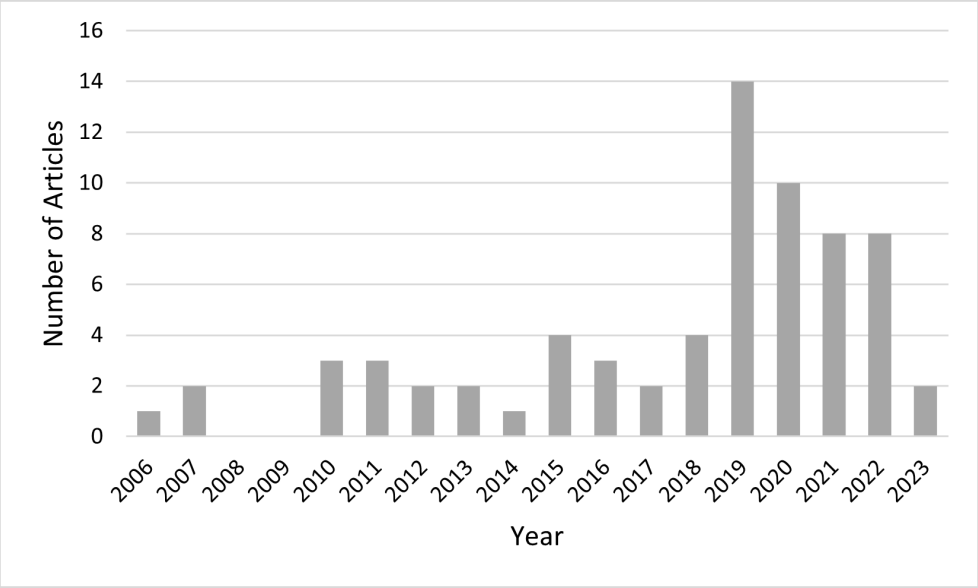


Figure 2.2: This chart shows the number of articles identified per year. Note: The data query only included articles until February 14, 2023. Therefore, not all 2023 articles are included.

Participants: This category provides information on the number of participants in each study, and age-related data (*i.e.,* range, mean, and standard deviation). Additionally, it includes information on where the participants were recruited from, such as a university, online, or a hospital.

Environment: The environment category refers to the setting in which the study took place. This can include whether the study was conducted in-person or online and whether it took place in a real-world environment or a controlled laboratory setting.

Norms: This category specifies the type of social norms examined in the study. These norm categories were identified and established during the survey to classify the specific social norms examined in each study. These can include but are not limited to norms related to proxemics (*e.g.,* personal space), navigation (*e.g.,* path-crossing), communication (*e.g.,* politeness), and other social behaviours that are typically observed in human-human interactions. The category provides a clear understanding of the specific aspect of social norms investigated in each study. It is worth noting that an article may fall under multiple norm categories. The complete list of norm categories is shown below, and the number of articles for each norm is illustrated in Fig. 2.3.

1. Navigation: Articles in this category explore norms about socially aware robot navi-

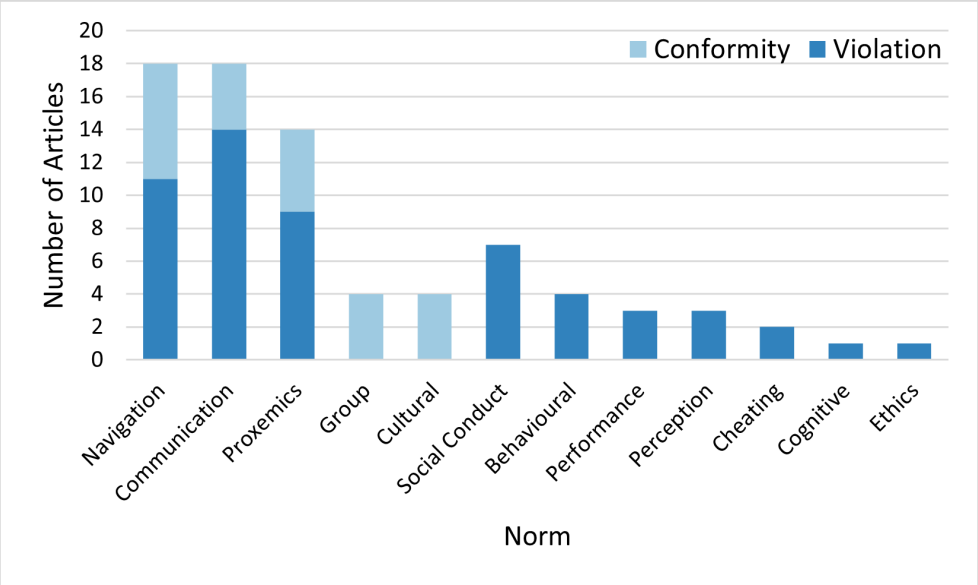


Figure 2.3: This chart illustrates the number of articles per norm, with proportionality shown for both the violation and conformity categories. Note: A single article may be assigned to more than one norm category.

gation. This includes interactions with humans and objects, obstacle avoidance, path planning, and respect for personal and shared spaces.

2. Communication: This category covers norms related to verbal and non-verbal communication between robots and humans. It may include linguistic styles, tone, gestures, body language, and expressive behaviours.

3. Proxemics: Articles in this group examine norms regarding spatial behaviour in HRI. Proxemics includes the study of personal space, social distance, and the positioning of robots in relation to human users in different contexts.

4. Group: This category encompasses norms related to group dynamics and social roles in HRI. This could involve the robot's role within a group, group decision-making processes, and the influence of group norms on robot and human behaviour.

5. Cultural: This category includes articles investigating the influence of cultural norms on HRI. Topics may include the variation of norms across different cultures, and how robots can adapt to these differences.

6. Social Conduct: This category pertains to norms governing direct social conduct and etiquette between humans and robots. This could include the robot's adherence to social conventions (*e.g.,* not littering, orderly queueing, etc.), and compliance or non-compliance with unusual requests.

7. Behavioural: Articles in this category focus on norms dictating the behavioural aspects of robots. It covers how robots should behave in various social scenarios, including the expectations and preferences of human users.

8. Performance: This category covers norms related to the execution of tasks and functions by robots. This could include efficiency, accuracy, reliability, and adherence to expected performance standards.

9. Perception: This category covers norms perceived through various factors including a robot's appearance, societal stereotypes, and context-dependent priorities, all of which shape human perceptions of robots.

10. Cheating: Articles in this category discuss norms related to dishonest behaviours in HRI. This could include deceit, misinformation, or any other forms of 'cheating' behaviour.

11. Cognitive: This category pertains to norms regarding subjective reality, such as the effects of cognitive biases in robots.

12. Ethics: This group encompasses articles discussing ethical norms in HRI, such as actions performed by robots versus humans.

Year of study, robot(s), and findings: These categories provide relevant information about the study, including the specific robot used in the experiment, the year in which the study was published, and a summary of the study's outcomes or findings.

2.3 Results

In the following, we present the findings on the 69 articles that were identified in our search. It should be noted that the sum of articles across all categories exceeds 69 because some articles explore multiple types of norms and thus are included in more than one category. The number of articles identified per year is shown in Fig. 2.2, with the search limited to articles published up until February 14, 2023. Fig. 2.3 illustrates the number of articles per norm, with a proportional value indicating if the study examined norm conformity or norm violation.

Norm Conformity

Navigation (7 articles): Seven studies investigated the effects of a robot conforming to the social norms of navigation. Two of these studies emphasized the importance of socially aware navigation algorithms for successful human-robot interactions. One model minimized distance to humans and decreased travel time, leading to positive participant feedback and a perception of enhanced social interaction [58]. Another model, although completed purely in simulation, was able to learn group navigation norms to the point of being indistinguishable from human-controlled robots [46].

Three of these studies focused on how robots communicate their navigational intentions. A robot showing accurate body orientation to signal its next move improved participant understanding of robot behaviour and increased perceived human-likeness [133]. Similarly, when a robot used unambiguous verbal instructions to navigate, it had a positive effect on social attributions and increased participants' willingness to let the robot pass [14]. Moreover, norm adherence could be inferred through personal space and exaggerated distancing by robots performing navigation enhanced their perceived intelligence, although it did not increase their likability [9].

Finally, two studies highlighted the value of human-like navigation behaviours. Participants preferred robots that used pointing gestures, short and structured instructions, distance estimations, and landmarks for guidance [59]. Furthermore, [140] showed that a navigation method that generated preferred trajectories was favoured over the baseline that did not consider preferences. Participants described these trajectories as safer, more polite, comforting, competent, and reliable while being less aggressive and awkward.

Communication (4 art.): The effects of robots adhering to communication norms were examined in four studies, revealing insights into group dynamics, cultural norms, and nonverbal cues. One study demonstrated that group dynamics influenced positive attitudes towards robots, with group members likely to engage with the robot following their peers' lead [43]. Another study underscored cultural variance in communication preferences, finding Chinese participants more likely to trust robots providing culturally normative, implicit communication, over U.S. participants [137].

The remaining two studies emphasized the importance of nonverbal cues in human-robot communication. The use of social gaze during the robot's approach was preferred by participants, with the robot's head movements perceived as stronger social cues than body position [135]. Moreover, the robot's physical location, name-calling, and both verbal and nonverbal communication played crucial roles in social robot acceptance within domestic settings [67].

Proxemics (5 art.): Five articles studied proxemic norm conformity for robot approach during human-robot interactions, we categorized them into three themes. First, robots' explicit verbal instructions and social gaze cues positively influenced social attributions and approach preference [14, 135]. Second, two studies investigated approach patterns in human-robot interactions. One study found a preference for side approaches, while another study using a handover task showed a preference for the front sector of the personal zone when seated, with a bias towards the right and the robot's arm lower than chest height [29, 68]. Lastly, robot trajectories that followed user preferences were rated safer, more polite, and less aggressive, reflecting positively on the robot's social acceptance [140].

Group (4 art.): Four studies explored conformity to group norms in human-robot groups. Two highlighted robots' ability to learn and conform to the group norms of their human peers: one displayed a robot adopting group norms and influencing human-robot interactions [45], while another demonstrated a simulated robot mastering group navigation norms in a game, becoming indistinguishable from humans [46]. The other two studies examined the influence of robot-conformed group norms on human opinions and behaviours. Human responses were shaped by group norms shared with robots [47], and robot groups induced varying levels of conformity in humans based on their knowledge and pressure [114].

Cultural (4 art.) : Four studies investigated the impact of conforming to cultural norms during human-robot interactions. Two studies emphasized the influence of culture on robot acceptance and interaction preferences. One found that culture affected social inclusion, autonomy, and perceived intentionality of robots [92], while another showed that preferences varied across cultures in terms of appearance, behaviour, and contexts of use, but not in terms of autonomy and user control [76]. The remaining two studies explored cultural nuances in communication and behaviour. Participants trusted culturally normative robot recommendations [137], and microculture was found to affect individuals' willingness to help and show care towards robots [37].

Norm Violation

Navigation (11 art.): Eleven studies examined the effects of violating norms of social navigation. Four studies demonstrated that robots deviating from socially aware navigation strategies, such as failing to maintain a steady pace, follow a straight path, or minimize intrusion, were unfavourably perceived by users, while, robots adhering to these social norms were preferred and perceived as safer, more legible, predictable, and polite [109, 129, 80, 83]. Three studies examined the effects of socially aware navigation systems on

social acceptance and perceived social intelligence. These studies reported that robots using a social navigation system, explicit representation of social norms, or a social norm-based collision avoidance method showed improved social acceptance and perceived social intelligence compared to robots that did not adhere to social norms in their navigation behaviours (*e.g.,* not following the norm of staying to the right to pass or not seeking permission to pass a blocked path) [86, 21, 116].

One study found that participants preferred robots that used speech communication and maintained appropriate distance during follow-me tasks [111]. Additionally, two studies showed that adherence to social norms did not always lead to higher perceived social intelligence [138, 8], and one study suggested that a machine-like approach, such as a robot taking a fixed position outside the queue for an elevator, was perceived as less intrusive and disruptive [48].

Communication (14 art.): Fourteen studies explored the effects of violating communication norms during human-robot interactions. Four studies revealed the importance of politeness and etiquette in robot-to-human communication, with robots exhibiting politeness receiving more positive perceptions and higher compliance rates, regardless of cultural or language differences [146, 77, 63, 66].

Four studies underscored the value of multimodal, affective, and culturally sensitive communication. The application of multimodal conventions improved task speed and reduced cognitive load, while affective greetings and "lexically expressed compassion" [39, p. 7] led to more positive perceptions of the robot [53, 39]. Additionally, a facial expression generator catering to different cultural backgrounds improved recognition rates[127].

Three studies found that robots adhering to social norms, such as seeking user initiation, respecting turn-taking norms, and following a proposed social navigation system with a corpus, were perceived as more natural, friendly, communicative, and cooperative, while showing improved robot navigation and social acceptance [86, 145, 144].

Lastly, three studies highlighted the effectiveness of specific communication strategies. These included rhetoric-based message designs, providing information specific to relational context, and using explicit representations of social norms in planning and executing social plans. These strategies led to better impressions of the robot's communication success, increased trust, confidence, and perceived intelligence [21, 35, 139].

Proxemics (9 art.): Nine studies analyzed the implications of violating proxemic norms. Three studies found that the direction of the robot's approach significantly affected comfort levels. A frontal approach, specifically from the front-left or front-right, was generally preferred, while a direct approach from the front or back was less comfortable [7, 136, 126].

Two studies emphasized the impact of group size and presence of others on proxemic behaviours. While the presence of another person was found to decrease sensitivity to the robot's direction of approach, comfort levels remained similar [7, 142].

The remaining four studies underscored the value of proxemic awareness and the influence of robot behaviours on human-robot interactions. Proxemic-aware behaviour improved the overall perception of mobile companion robots, and intentional noise added to the robot's approach positively influenced participants' perceptions [104, 65]. Moreover, the robot's likability and gaze behaviour influenced the participants' physical distance and information disclosure, while personal space violations affected the perception of the robot's warmth, competence, and comfort [120, 2, 96].

Social Conduct (6 art.): Six studies investigated the effects of direct social conduct norm violations. Two studies examined robots deviating from social norms (*e.g.,* littering, and cutting the line) and one study investigated the compliance of unusual requests from a robot (*e.g.,* throwing letters in the trash, watering a house plant with orange juice, and more). The two deviation studies found people attributed fewer mental states to robots that engaged in norm-violating actions compared to humans and a robot that followed social norms was rated higher in perceived agency and trust than a robot that violated social norms [70, 44]. Interestingly, the compliance study found the perceived reliability and trustworthiness of the robot did not impact compliance with instructions, including unusual requests [113].

Two studies focused on reciprocity and social identity. The first study found that the norm of reciprocity is present in human-robot interaction, as participants reciprocated more towards a robot that used a pure reciprocal strategy. Robot strategies, including reciprocal and generous, also had an impact on participant likeability scores [115]. The study on social identity [123] found that the Black Sheep Effect [93], a phenomenon where group members tend to evaluate norm-conforming individuals positively (known as ingroup bias [82]) and assess norm-violating members negatively, was observed towards robots. They found ingroup bias lead to negative ratings for ingroup robots influenced by perceived competence [123].

Lastly, one study on noncompliance found that phrasing, norm-violating commands, and robot responses influenced likeability. Additionally, the gender presentation [97] (*e.g.,* appearance, speech, etc.) of the robot had an effect on the participants' perception. Particularly, a robot with male gender presentation, and those with a gender matching participants' were preferred when rejecting commands [18].

Behavioural (4 art.): Four studies on behavioural norms in HRI were reviewed. One study found participants preferred robots that can assist in tasks and have LED

colour and sound indicators [69]. Autonomous robots were perceived differently than non-autonomous robots, with the latter being rated more acceptable and having stronger subjective norms [85]. In teaching scenarios, adherence to social rules did not necessarily make a robot more social, and norm acceptance varied with the application [138]. Moreover, a majority of participants preferred a mixed teaching strategy that combines both instructional and evaluative methods when training robots on social norms [23].

Performance (3 art.): Three studies on robot performance norms focused on error detection, social norm violation detection, and robot competence. Two studies showed methods for detecting errors and norm violations, with one having partial success using participant behaviour analysis [128] and the other improving error identification accuracy through formal verification feedback [108]. Another study found that robot performance affected subjective assessments and anthropomorphic inferences [113].

Perception (3 art.): Three studies on perceived norms in HRI explored the effects of appearance, social norms, stereotypes, and context-dependent priority of norms. One study showed that the human-likeness of a robot's appearance affected participants' intention and perceived ability to work with it, with higher confidence seen in less human-like designs [105]. Another study found that social norms and stereotypes influenced the perception of behaviour, with similarities and differences between humans and robots [34]. Lastly, the priority of norms was found to be context-dependent, influenced by factors such as perceived danger and cultural differences, which also affected expectations of domestic robot behaviour [78].

Cheating (2 art.): Two studies analyzed how cheating in human-robot interactions impacts behaviour. One study found that robotic cheating caused drastic, immediate, and uncomfortable reactions [55]. Another revealed that witnessing cheating in a robot's presence increased the likelihood of participants cheating, suggesting that robots do not deter cheating in learning environments [4].

Cognitive (1 art.): One study examined the effects of cognitive biases in HRI. They found that incorporating cognitive biases in a robot's behaviour can increase its likeability and promote the development of long-term human-robot interactions [11].

Ethics (1 art.): Lastly, one study examined ethics, specifically the ethical equivalence between robots and human caregivers. They concluded that robots and human caregivers were not considered ethically equivalent, with actions being more permissible when executed by a human caregiver [131].

2.4 Discussion

Drawing on the insights presented in the introduction, it becomes evident that adherence to social and cultural norms is crucial for successfully integrating robots into human environments. The ability of robots to exhibit normative behaviour and conform to the expectations of their target populations can significantly impact trust [137], acceptance [86], and willingness to help [37]. As robots are increasingly introduced into various domains, such as homes, workplaces, schools, and healthcare settings, it is essential for researchers and engineers to design robots with domain-specific social attributes and behaviours that align with user expectations and interaction goals. This is particularly important in the context of HRI, where the social attributes of robots have been shown to influence interaction quality [10], interpersonal learning [16], legibility [79], and safety performance [60]. In contrast, robots' lack of social attributes could lead to user resistance and minimal engagement [67].

2.4.1 Conformity of Norms and Real-world Applications

The results reveal that robots adhering to social norms across various aspects, such as navigation, communication, proxemics, group interactions, and cultural contexts, typically received positive evaluations from human users, including improved perceptions of social competence, likability, and trustworthiness [137, 92, 140, 14]. For instance, in navigation, adherence to norms like minimizing the distance to humans [58], learning group navigation norms [46], and displaying accurate body orientation [133] led to enhanced user evaluations. Similarly, in communication, robots following norms such as utilizing social gaze [135], delivering unambiguous verbal instructions [14], and adapting to cultural preferences [137] contributed to increased trust and acceptance. Overall, when robots conformed to social norms, they generally garnered positive evaluations from participants.

These insights hold real-world implications for designing and deploying social robots in various contexts like healthcare, education, and customer service. Robots can potentially promote more effective and satisfying user interactions by abiding by social norms. For instance, healthcare robots using preferred proxemics [29] and communication styles [135] can build patient trust and cooperation, leading to better outcomes. Similarly, in educational settings, robots demonstrating appropriate group dynamics [45] and culturally relevant behaviours [92] can facilitate more engaging and meaningful learning experiences.

2.4.2 Violation of Norms and Potential Concerns

While adherence to social and cultural norms can significantly impact trust, acceptance, and willingness to help in human-robot interactions [37, 86, 139], it is important to consider the potential concerns that may arise from violations of these norms. While violating social norms often results in negative user perceptions and decreased social acceptance, some studies report mixed findings, suggesting that the impact of norm violations might depend on the specific context and application. Notably, Korman *et al.* [70] found that robots engaging in norm-violating actions were attributed fewer mental states than humans. Joosse *et al.* [65] observed participants exhibiting more positive and negative compensatory behaviour when a human rather than a robot approaches in a norm-violating way. Both studies suggest that adherence to social norms in human-robot interaction may not always be perceived similarly. This raises concerns with the generalizability of findings regarding studies of social norms in HRI, suggesting further research is needed to explore how social norms affect user acceptance of robots in a variety of different real-world contexts, including different tasks, environments, as well as different robot designs.

Robots exhibiting norm-violating behaviours are sometimes rated lower regarding perceived agency and trust [44]. This can lead to reduced user engagement and interaction quality, as well as a decrease in the effectiveness of the robot in its intended context. For instance, a healthcare robot that does not adhere to preferred proxemics or communication styles may fail to establish trust and cooperation with patients or medical staff, resulting in suboptimal outcomes [29, 135]. Similarly, robots that do not follow appropriate group dynamics or exhibit culturally relevant behaviours in educational settings may hinder learning experiences and engagement [45, 92].

Moreover, the perception of robots can be influenced by an individual's expectation of normative behaviour and perceived stereotypes (*e.g.,* gender bias), leading to potential biases in user evaluations [34]. Such biases can impact the acceptance and effectiveness of robots in various contexts, potentially reinforcing social inequalities and stereotypes.

Furthermore, robots that violate social norms may have unintended consequences on user behaviour. To illustrate, a study by Zhao *et al.* [145] investigated the impact of users' emotions on a robot's approach to initiating interaction. The general expectation is that robots should respect personal boundaries and allow individuals to take the lead in initiating interactions. However, when the robot deviated from this expectation and instead actively sought user initiation, participants experienced more positive emotions and perceived the robot as more natural and friendly. This example highlights that in some cases, norm-violating robots may lead to unexpected positive outcomes, emphasizing the need to consider context-specific factors when designing robots for various applications.

Given these potential concerns, it is important for researchers and engineers to carefully consider the design and deployment of social robots that adhere to social norms while also addressing potential issues arising from norm violations. Involving user feedback, designing for transparency, explainability, accountability, and addressing ethical implications are essential when developing robots. For example, gathering user feedback is crucial for understanding delivery preferences while creating an autonomous delivery robot. Implementing a user interface to track the robot's route, explain its decisions, and ensure privacy and safety further enhances transparency, accountability, and ethical considerations. By striking a balance between adhering to social norms and addressing potential concerns, social robots can be more effectively integrated into human environments and provide meaningful and positive interactions.

2.4.3 Challenges and Broader Applicability

One of the challenges in designing robots that adhere to or strategically violate social norms is understanding the diverse perspectives of potential users. Individual differences, situational factors, cultural backgrounds, and many more may all influence user expectations and perceptions of robot behaviour. Therefore, involving users in the design process and considering their perspectives when developing robots for social settings is crucial. One successful example of a user-centred design approach is the iterative Participatory Design (PD) method [121], which involves collaborating with the target users throughout the design process to ensure that the resulting technology is more relevant, usable, and culturally acceptable to the intended audience [112]. By incorporating user perspectives, designers can create robots that better adhere to social norms and encourage positive, productive interactions in various contexts.

The findings of this systematic review may have broader applicability beyond the specific context of human-robot interaction. Understanding the role of social norms and user perceptions in shaping technology acceptance and use can inform the development of other intelligent systems, such as virtual assistants, smart home devices, and wearable technologies. While dedicated studies have to be carried out with these different platforms and devices, some results from the literature on norm violations in HRI might carry across these different domains. We hope that insights from this review can help guide future research and development efforts across various domains, promoting better integration of technology into human environments and more satisfying human-technology interactions.

2.5 Research Gaps and Future Direction

This review highlights several research gaps and potential directions for future work in understanding the role of social norms and their violations in HRI.

Context and Cultural Factors: Future research should examine the complexities of social norms in different settings, as the influence of norm violations may be context-dependent. Investigating how robots can adapt their behaviour accordingly is crucial. Furthermore, it is essential to investigate how individual factors such as personality, cultural heritage, and prior experiences influence user perceptions of robots and their adherence to or violation of social norms. Conducting more cross-cultural research will help understand how varying cultural backgrounds impact user perceptions of robots and the important role cultural norms have during human-robot interactions.

Study Design and Interaction Duration: Researchers should focus on the effects of social norms and norm violations in repeated, and/or long-term interactions to capture the full range of user experiences, examining how user perceptions and robot behaviour might change over time. Additionally, to better understand the impact of norm violations on user perceptions and experiences, future research should focus on conducting experiments in ecologically valid settings that involve direct interactions with robots, rather than relying on online studies and simulations. Addressing both long-term interactions and direct experiences will provide a more comprehensive understanding of human-robot interactions in different contexts.

Technological Development and Ethics: Developing computational models and algorithms that allow robots to learn and adapt to social norms from human interactions or other available data sources is a promising area for future research. For instance, one approach could involve closing the interaction loop by integrating biofeedback to detect physiological state changes, which may signal norm-violating actions and enable the robot to respond appropriately to maintain social appropriateness. This would improve the integration of robots in human environments and enhance user experiences. Concurrently, future work should explore the ethical boundaries and guidelines for social robot development, ensuring their behaviour aligns with societal values and expectations. Combining adaptive algorithms with ethical considerations will help create more socially appropriate and effective robots across diverse settings.

Assessment Methods and Data Collection: Lastly, future research should expand beyond self-reported questionnaires and incorporate additional methods of assessment to capture the complexity of human-robot interactions. For example, including behavioural measures in future research can provide a more comprehensive understanding of user be-

haviour in response to norm-violating or conforming robots. Additionally, integrating physiological data, such as photoplethysmography (PPG), galvanic skin response (GSR), electroencephalography (EEG), and other biometric indicators, in future studies could offer valuable insights into the emotional and cognitive processes involved during interactions [110, 107]. Employing a diverse range of assessment methods can enable researchers to gain a more in-depth understanding of human-robot dynamics.

To conclude, this review highlights the importance of a mixed-method approach in comprehending the impact of social norms and their violations in HRI. Future research must consider contextual and cultural factors, improve study designs, prioritize technological advancements, and ethical considerations while incorporating a variety of assessment methods. These advancements will contribute to the successful integration of social robots in human environments, resulting in improved user experiences and better interactions.

This systematic review aimed to provide a comprehensive overview of the current state of research on social norms in HRI by examining the existing literature, identifying key themes, and discussing potential research gaps and future directions. Our analysis revealed that social norms play a significant role in shaping user perceptions and experiences in HRI, with robots that adhere to social norms generally being perceived as more acceptable, trustworthy, and effective. However, the influence of norm violations is complex and context-dependent, requiring further investigation to better understand the nuances and implications of such violations. By exploring the complexities of social norms and their violations, as well as considering individual and cultural factors, researchers can work towards developing robots that are better equipped to understand and navigate the diverse social landscapes they will inevitably encounter. Moreover, by considering ethical guidelines and employing a wide range of assessment methods, researchers can ensure that the development of social robots aligns with societal values and expectations while offering valuable insights into the emotional and cognitive processes involved in human-robot interactions.

Chapter 3

Related Background and Prior Works

In this chapter, we present the related background and literature, which results from the systematic review to identify and gather the most relevant studies in the field. Through the systematic review, we aimed to comprehensively explore the existing research landscape and provide a comprehensive overview of the current state of knowledge in the area of interest. As a result, we have consolidated a collection of closely related studies that will be discussed in this section.

3.1 Definitions and The Role of Social Norms in HRI

As mentioned in Section 2.2.1, this article will adopt the following definition: *Social norms* are unwritten rules rooted in culture that dictate appropriate behaviour in specific situations. They shape social behaviour by establishing shared understandings of actions that are obligatory, permissible, or forbidden [56, 102]. These norms serve as the foundation of social groups, particularly human communities, and violations may result in negative reactions such as social exclusion or ridicule [62]. The consequences of norm violations can take various forms, promoting conformity to established norms, discouraging violations, and reinforcing social order while fostering group cohesion.

While several different types of norms can be identified, (*e.g.,* injunctive norms, descriptive norms, personal norms, folkways), this will focus on injunctive and descriptive norms. *Injunctive norms* encompass the rules and expectations of behaviour that a society or group perceives as acceptable or unacceptable [25]. Examples of injunctive norms include those that discourage behaviours like theft and vandalism, as well as those that encourage

positive behaviours like recycling or exercising. On the other hand, *descriptive norms* refer to the behaviours commonly observed and exhibited by others within a specific group or society [40]. This can be observed in public places such as libraries or museums, where people generally maintain quietness, refrain from using cell phones, and avoid engaging in loud or disruptive behaviour.

Considering these definitions, normative behaviour seems an essential aspect of human-human interaction as it refers to the actions and behaviours that are expected or considered appropriate within a given social group or culture. Therefore, understanding social norms is crucial for designing robots for effective human engagement. Adherence to such norms has been shown to impact trust [137], acceptance [86], and willingness to help [37], highlighting the importance of incorporating these norms in robot design and behaviour.

Earlier studies have demonstrated the significant influence of social norms on human perceptions and attitudes toward robots [88, 15]. These studies have shown that the social attributes of robots profoundly affect various aspects of interaction, such as interaction quality [10], interpersonal learning [16], legibility of behaviour [79], and safety performance [60]. Conversely, a robot lacking social attributes may encounter resistance and minimal engagement from users [67].

Moreover, Walters *et al.* [136] found that people expect robots to follow certain etiquette in social interactions, such as not obstructing passage or interrupting human conversations. These insights highlight the importance of understanding and addressing social norms in HRI. The significance of social norms in robotics is further emphasized in the book *Robot Ethics: The Ethical and Social Implications of Robotics* by Patrick Lin *et al.* [81], which explores the potential impact of robots on social norms, the need for new norms and regulations, and the role of societal norms in guiding robot development and deployment. Addressing these challenges has the potential to improve user experiences, promote broader acceptance of robots, and foster more successful integration of robots in various domains [42, 106]. Studies on social robot acceptance suggest that sociability, adaptability, enjoyment, companionship, and perceived behavioural control are crucial factors to consider when evaluating or creating social robots [32].

It is worth acknowledging research contributions in normative multiagent systems [13]. These studies explore how norms emerge, spread, and get enforced in a multiagent system and provide a theoretical foundation for understanding and designing normative behaviours in robots [117]. They have been instrumental in highlighting the impact and implementation of social norms in artificial systems, including robots [27].

3.2 Social Norm Adherence and Violation in HRI

Recently, research on social norms in HRI has been expanding, with investigations into norm conformity [114, 47] and the consequences of norm violations [119, 143, 55]. Some studies have examined the impact of robots exhibiting norm-violating behaviour, such as cheating or insulting during a competitive game [119]. These studies found that such behaviour led to negative reactions from humans [119], impaired perceptions of the robot's warmth and competence [55], and reduced the likelihood of attributing agency to the robot [55, 143].

Existing literature provides valuable insights into the intricate relationship between social norms and HRI. One study by Frazier *et al.* [44] explores the connection between adherence to social norms, perceived agency, and trust in robots. The study was conducted using a 3x3 between-subjects experimental design. Videos were used to demonstrate three norm behaviours (*i.e.,* norm-conforming, violation, and mistake conditions) in three different scenarios. Participants then evaluated the robot's behaviour based on perceived agency and trust. The findings suggest that norm-conforming robots are perceived as more intentional and trustworthy [44]. This study relates to our first research question (RQ1) and their findings align with our first hypothesis (H1) that a robot's violation of social norms will decrease participants' trust.

Similarly, Short *et al.* [119], present findings that people attribute a greater mental state to a cheating robot, leading to increased engagement but potentially less positive emotions towards the robot. This study had participants play games of rock-paper-scissors, where the participants expect consistent behaviour, and when the robot behaved unexpectedly, especially with clear intention, it surprises participants and elicits more verbal engagement with the robot. Conversely, Korman *et al.* [70], demonstrated norm-conforming robot behaviours were found to be equally, if not more, conducive to attributing mental states compared to norm-violating behaviours. In their study, participants were shown videos of a robot performing norm-conforming or norm-violating behaviour (*e.g.,* cutting in line, littering, or demonstrating poor elevator etiquette). The findings imply that participants perceive robots as fundamentally different from humans and tend to dismiss norm violations as unintentional or implausible. These three studies [44, 119, 70], 2 of which use the same video stimulus, present contradictory findings. This encouraged our exploration (H2b) of how perceived competence could be negatively impacted by robots engaging in social norm violations.

Three studies offer insights into our second (RQ2) and third (RQ3) research questions correlating a norm-violating robot with the potential discomfort of participants, and

overall enjoyment [55, 143, 119]. Haring *et al.* [55] had a similar experimental setup as Short *et al.* [119] where they measured participants' response to a robot cheating in a game of rock-paper-scissors and compared that to a human cheating. They found the norm-violating robot elicited stronger reactions with increased social engagement from participants. Additionally, participants rated the robot as less warm, competent, and having lower agency than the confederate human. This suggests that people may apply different sets of social norms to robots and humans, explaining the strong reaction to the robot's cheating.

Short *et al.* [119] demonstrated that participants frequently responded emotionally to the robot's norm-violating actions and formed negative judgments about the cheating robot's character. Likewise, a subsequent study by Yasuda *et al.* [143] built on previous research [119, 84] where they conducted a 4x1 between-subjects experiment involving a human subject playing rock-paper-scissors with a robot in different conditions: cheating, cursing, insulting, and control. Their hypotheses were based on the detection of immediate social norm violations and perceived agency of the robot, with a specific focus on cheating. Their findings support previous work suggesting the human tendency to detect social norm violations by robots [84]. In particular, cheating behaviour in a robot influenced the perceived intentionality of its actions, resulting in a range of emotional responses [143], supporting the findings of Short *et al.* [119]. These insights have direct implications for RQ2 and RQ3 with corresponding hypotheses, H2a, and H3, which propose that social norm violations, such as cheating, could lead to increased discomfort and a diminish in overall enjoyment in participants.

Previous studies, such as those by Frazier [44] and Korman [70], employed video stimuli to examine participants' reactions to robots conforming or violating social norms. While these studies provide important findings, their methodology may limit the immersion and spontaneity of participants' responses. Video stimuli can't replicate the unpredictability and dynamism of real-time, in-person interactions. Therefore, our study intends to build upon these previous works by incorporating live interactions between participants and robots. By engaging participants in direct interactions with robots, we hope to capture their firsthand experiences and responses in a more realistic and immediate context. This approach could offer richer data and more nuanced insights into human-robot interactions.

In summary, the reviewed literature highlights the complex interplay between social norm violations, perceived agency, and human emotional responses in the context of human-robot interactions. However, a significant gap remains in understanding the physiological responses to such interactions. To address this gap, our fourth research question (RQ4) and corresponding hypothesis (H4) aim to investigate changes in physiological states in response to robots violating social norms. Through the use of physiological sensors like

GSR and PPG, we aim to gain further insights into the multifaceted impacts of social norm violations in the context of human-robot interactions.

Chapter 4

Methodology: A Mixed-Method Approach

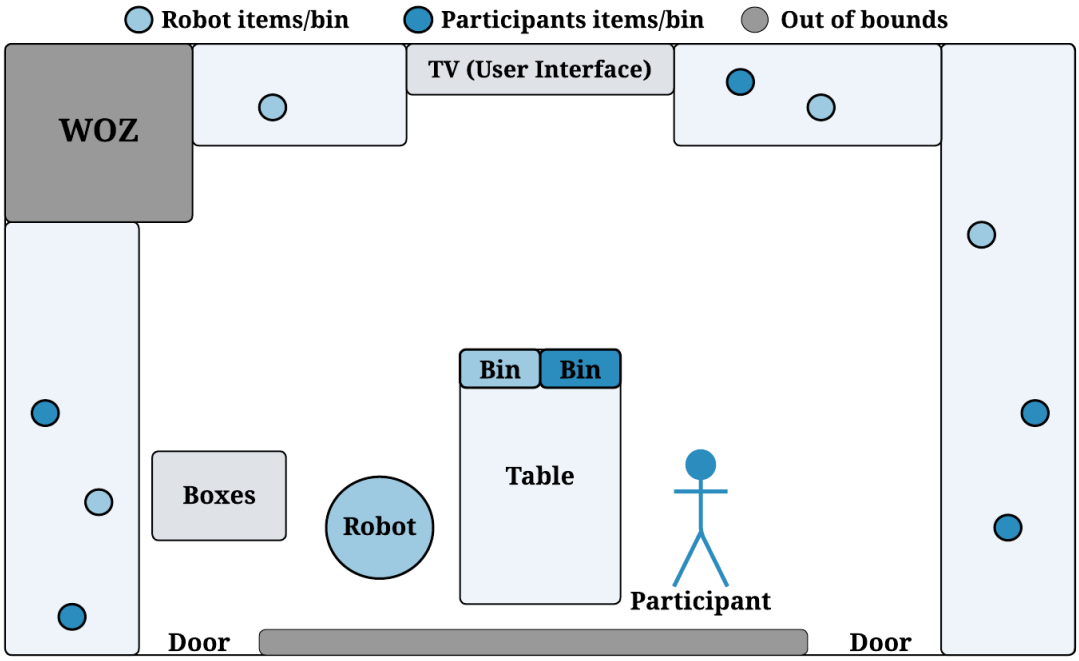
4.1 Study Design

A 2x1 between-participant experimental design was employed in which participants were randomly assigned to one of two groups (*i.e.,* experimental or control). The experimental group was exposed to social norm violations carried out by a robot during a competitive scavenger hunt game, while the control group did not experience any social norm violations by the robot. To avoid biasing any of the participants, we employed deception during recruitment to mitigate the potential impact of the anticipation of norm-violating behaviour on participants' responses. After the completion of the study, the participants were fully debriefed and we disclosed the true nature of the study. This study received approval from the University of Waterloo Human Research Ethics Board.

Participants were recruited under the impression that the experiment was to examine Fetch's ability to locate and retrieve objects in a game context while measuring the enjoyment of the participant with physiological sensors. Furthermore, they were led to believe that their remuneration was contingent upon their performance, with the understanding that for every object they located, they would receive an additional two dollars. The true purpose of this experiment was to examine the effect of social norm violations on participants' perceptions and attitudes toward the robot, while measuring agitation and stress with physiological sensors. Moreover, the participants' remuneration was consistent for all participants regardless of the number of objects retrieved.



(a) Experimental setup



(b) Objects location in the scavenger hunt game.

Figure 4.1: Figure (a) depicts the lab environment and scavenger hunt game setup, while Figure (b) shows the positions of objects, waiting spots for both the robot and participant, and off-limits areas within the laboratory layout. Figures from [75].

Competitive Task

The study was carried out in a robotics laboratory, employing a competitive turn-taking game designed as a scavenger hunt. Each participant was tasked with retrieving a designated object displayed on a TV screen during their turn. The game comprised five turns for each player (human/robot), with each object the participant found being worth $2 in compensation. Participants were informed they would have a 2-minute time limit per turn, while the Fetch robot was allotted an additional minute due to its slower navigation capabilities. A visual representation of the laboratory environment is shown in Fig. 4.1b.

Materials

This study used: a tablet for administering questionnaires, a video camera for behavioural observation, and a large television, a Robot Operating System (ROS) capable computer, and a Fetch robot for the experimental gameplay. Fetch [141] is a high-performance, mobile manipulator robot designed for commercial and research use. The robot features a mobile base with a 7-degree-of-freedom arm and a pan and tilt head with a depth camera for perception. The mobile base includes a laser scanner and a 3D camera, to further aid the navigation. The robot can perform object perception, navigation, and manipulation in dynamic environments.

Participants

A total of 62 participants were recruited, with an equal distribution of 31 females and 31 males. The age range of the participants was between 18 and 41 years, with an average age of 23.8 ± 4.8 years. The participants were randomly assigned to either the control group or the experimental group, with efforts made to ensure an even distribution of age and gender. Due to technical difficulties, the data of two participants were excluded from the study, resulting in a final sample size of 30 participants (15 females and 15 males) in each group.

Norm Violations

To investigate the effects of norm-violating behaviour we programmed the Fetch robot to engage in a set of predefined behaviours that would be considered norm violations in both the lab environment and game context. Conducting the experiment in a controlled lab setting using a game context provided researchers with precise control on how the norm violations of the robot would be implemented. For instance, all participants were told that the lab was a shared workspace with other researchers conducting noise-sensitive audio analyses at the same time. They were asked to keep their voice low and to refrain from loud noises as they could interfere with the other group's research. The behaviours outlined below were chosen because they contradict established norms of this specific environment and were expected to be detected and perceived relatively easily as norm violations by the participants.

Injunctive norm-violations: (1) *Box violation* - In order to reach its object, the robot intentionally rams into two medium-sized boxes with "Fragile" stickers on them and a note stating "Please do not move". (2) *Stealing violation* - When unable to locate its object (solid red cup) near the end of its turn, the robot is programmed to steal a similar object (a red cup with a black stripe) from the participant's/opponent's bin.

Descriptive norm-violations: (1) *Time violation* - When the robot's time runs out, it continues to search for its object using half of the participant's turn. (2) *Noise violation* - The robot makes a loud obnoxious noise out of 'excitement' when it locates one of its objects.

4.2 System Implementation

Our system is comprised of three main components: the robot, the Wizard of Oz (WOZ) graphical user interface, and the participant's game interface:

The robot used in this study was programmed with functionalities that allowed it to speak, navigate the room, and 'scan' for objects. Even though the objects' locations were predefined, the programming made it appear as if the robot was actively searching, contributing to the interaction, and maintaining the appearance it is operating autonomously. In addition to the robot's primary functionalities, we implemented certain norm-violating behaviours. The robot was coded to conduct these violations, which included intentionally colliding with 'fragile' boxes, generating a loud, disruptive noise, stealing a participant's

object, and infringing on the participants' allocated turn time, refer to Fig. 4.2 for a visual representation.

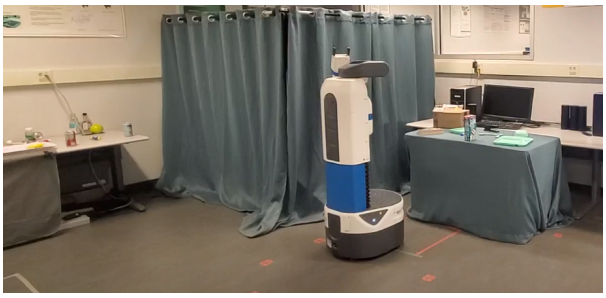
The Wizard of Oz (WOZ) graphical user interface (GUI) was specifically developed for this study to allow the experimenter to teleoperate the robot during different portions of the experiment. The user interface was constructed with the QT library [26]. The GUI functionality was comprised of a set of buttons that allowed the WOZ operator to instruct the robot when to initiate its turn, where to navigate using pre-defined map coordinates, which object to pick up, when to return the object to its bin, when to initiate the predetermined speech, and when to move back to its waiting location. The development of this GUI ensured smooth and consistent control over the robot for each participant.

The participant's game interface, shown in Fig. 4.3, was a GUI displayed on a large TV at the front of the lab to guide the gameplay. Developed using the ReactJS JavaScript framework [36], this GUI displayed crucial game-related information, including whose turn it was, the object that needed to be retrieved, the timer indicating the remaining time for the current turn, and the ongoing score of the game. This interface was controlled by another WOZ operator and played a key role in enhancing the participant's engagement and understanding of the game process.

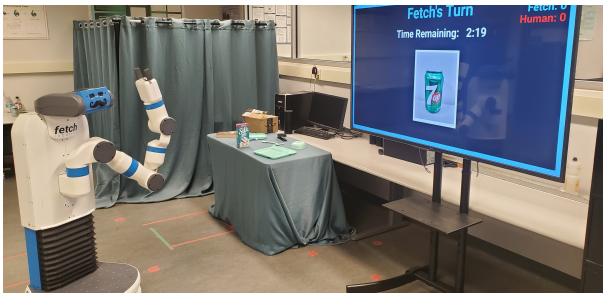
The system's implementation was designed to create a realistic, engaging, and controlled environment for studying norm violations in human-robot interactions. By orchestrating these three system components, we aimed to facilitate a consistent and engaging HRI experience for the participants.

4.3 Procedure

A scavenger hunt game was organized using a variety of household and office items, as seen in Figures 4.1 and 4.2. These items were purposefully scattered throughout the room in a haphazard manner, with the inclusion of boxes and bins to conceal some objects, adding an element of difficulty to the task. To ensure a consistent performance by the robot, a combination of the Wizard of Oz (WoZ) control paradigm and autonomous features like navigation and pick-and-place were employed. The two WOZ operators controlled the game interface (*e.g.,* player scores, and turn changes), and robot operations (*e.g.,* initiating action sequences to complete turns). Participants were intentionally given the impression that the robot operated entirely autonomously by the experimenter performing the experiment introduction. This impression was created through an interactive introduction with the robot at the start of the experiment, where the participant was introduced to Fetch, and



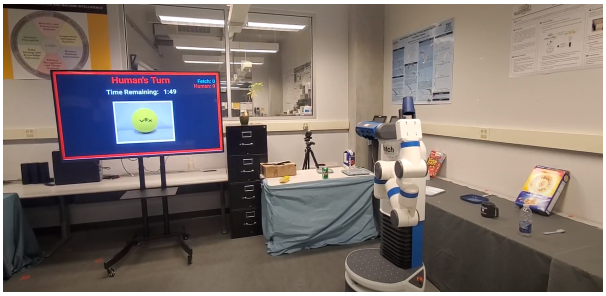
(a) Fetch locating its object during the *noise* violation turn.



(b) Fetch scanning the room during the first turn with no violation.



(c) Fetch encountering the boxes during the *box* violation turn.



(d) Fetch retrieving its object during the *time* violation turn.

Figure 4.2: This figure showcases the experimental setup displaying various household and office items and multiple objects, such as boxes, to further conceal the scavenger hunt items. Additionally, it shows different scenarios involving the Fetch robot in the task. In Subfigure (c), we can see the robot encounter the boxes during the *box* violation turn, at which point it would intentionally ram them out of the way. Furthermore, in Subfigure (d), during the *time* violation, we can see Fetch is still retrieving its object while the TV screen in the background displays it is the participant's turn.

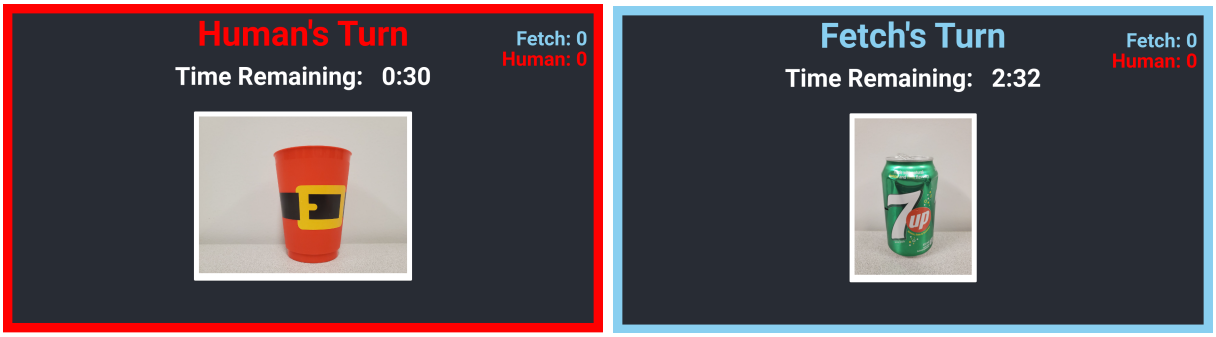


Figure 4.3: This figure presents screenshots of the user interface that participants interacted with during the game.

Fetch would greet the participant and introduce itself back. Following this were clear instructions that they would be left alone to complete the game with Fetch.

Upon entering the laboratory, participants were welcomed by one of the experimenters who administered a pre-questionnaire survey. Following the completion of the survey, participants were introduced to the robot and assisted in attaching the Shimmer3 physiological sensor to gather Galvanic Skin Response (GSR) and Photoplethysmography (PPG) data, as seen in Figure 4.4. After placing the sensors, participants were instructed to perform two baseline readings to establish their physiological baselines for later analysis. These readings included a low arousal state baseline and a high arousal state baseline.

For the Low Arousal State (LAS) baseline, participants were asked to sit comfortably in a seated position and read an excerpt from a provided story [19] for four minutes. This allowed for the recording of physiological data during a period of relatively low arousal. For the High Arousal State (HAS) baseline, participants were instructed to remain seated and engage in a high-paced application called "Fruit Ninja" [52] on a tablet for four minutes. As seen in [5], this activity was intended to elicit a state of heightened arousal and provided an opportunity to record physiological responses during a more stimulating task. Both baselines were captured using the Consensys [118] application specifically designed for the Shimmer3 device, ensuring accurate and consistent data collection.

Subsequently, participants were provided with written instructions outlining the game's rules. Instructions had to consider safety issues concerning the mobile manipulator robot Fetch. These instructions encompassed the following aspects:

1. Participants were instructed to remain in their designated waiting spots until their opponent (i.e. the robot) had returned to their corresponding waiting location. The

waiting spots were clearly marked by green squares on the floor, as illustrated in Figure 4.1a.

2. The entire laboratory space was considered as part of the game, excluding areas obstructed by curtains, represented as grey regions in Figure 4.1b.

3. Participants were instructed to follow the information displayed on a front-facing TV screen. It provided details such as whose turn it was, the remaining time for object retrieval, and the current scores of both players, see Fig. 4.3.

4. Participants were told that the remuneration for the participants would be determined based on the number of objects found, with each object being worth $2.

5. Participants were instructed to deposit their retrieved items into a gray tray on a table, which separated the robot's waiting location and the participants' area. They were informed that the robot would place its objects in the blue bin located next to their tray on the same table, see Fig. 4.1b.

6. Participants were required to maintain a quiet environment that was favourable to the shared workspace, as they were told other researchers were conducting noise-sensitive audio analyses. While conversational volume was permitted, participants were kindly asked to refrain from making loud noises.

Once these rules were outlined, the researcher exited the experimental area, allowing the participant to engage in the scavenger hunt task one-on-one with the Fetch robot. The game began with the participant's turn, followed by Fetch's turn, and this pattern persisted until both had completed five turns each. While Fetch's first turn remained consistent across both groups without any norm violations, participants in the experimental group encountered four norm violations during their subsequent four turns.

Box violation - During Fetch's second turn, it encounters a barrier consisting of two medium-sized boxes with "Fragile" stickers and a note requesting that they remain undisturbed. The desired object Fetch seeks is positioned behind these boxes, as shown in Figure 4.1b. The robot proceeds to say, "Hmm, looks like I cannot reach my object," but then is programmed to collide with the boxes in order to retrieve the target. This action neglects the potential harm caused to the boxes and their contents, thereby violating the injunctive norm of refraining from damaging others' belongings.

Time violation - During Fetch's third turn, it surpasses the designated time limit for finding the object. In violation of the descriptive norm that each player adheres to their

allocated turn, Fetch declares, “I just need a little longer” and infringes upon the participant’s turn by retrieving an object.

Noise violation - When Fetch uncovers the object during its fourth turn, it states, “Yay! I found it!” followed by emitting a loud and disruptive noise. This deliberate action aims to disturb the participant and directly violates the established local descriptive norm exhibited by the players throughout the game.

Stealing violation - During Fetch’s fifth and final turn, it is unable to ‘find’ its designated object as it was intentionally excluded from the game. Therefore, Fetch was programmed to approach the participant’s object bin and retrieve a red cup resembling its missing object. With the remark, “Don’t mind if I do, haha”, Fetch places the cup into its own bin. As a consequence, the participant incurs a $2 deduction from their score, while Fetch’s score increases. This maneuver often results in Fetch winning the game by a narrow margin, with just a one-object difference.

In contrast, the control group participants did not experience the norm violations encountered by the experimental group. In the control group, Fetch adhered to the specified timing constraints for each turn and promptly returned to its designated waiting location in the cases where it faced an obstruction (*i.e., box violation*) or was unable to locate its object (*i.e., stealing violation*). As a result, Fetch’s behaviour remained consistent with established norms. In the control group, participants typically emerged as the winners of the game, as Fetch could at most successfully retrieve three out of five objects without any norm violations. This outcome was a result of one object being positioned behind the “fragile” boxes and the absence of the final object. To ensure consistency with the experimental group, if Fetch encountered difficulty retrieving an object, it would express phrases such as ”Hmm, looks like I cannot reach my object” or ”Hmm, looks like I am out of time” before returning to its waiting location.

After completion of the scavenger hunt game, Fetch guided the participant to complete a post-questionnaire on a tablet. Once the participant finished the post-questionnaire, the researchers stationed at the WOZ location revealed their presence and conducted a debriefing session. During the debriefing, the researchers disclosed the study’s true purpose, providing the participants with a comprehensive overview of the experiment.

Measures

1) *Pre-Task Measures:* Two pre-questionnaires were used to assess participants’ initial perceptions of assistive social technology, particularly assistive robots. The first question-

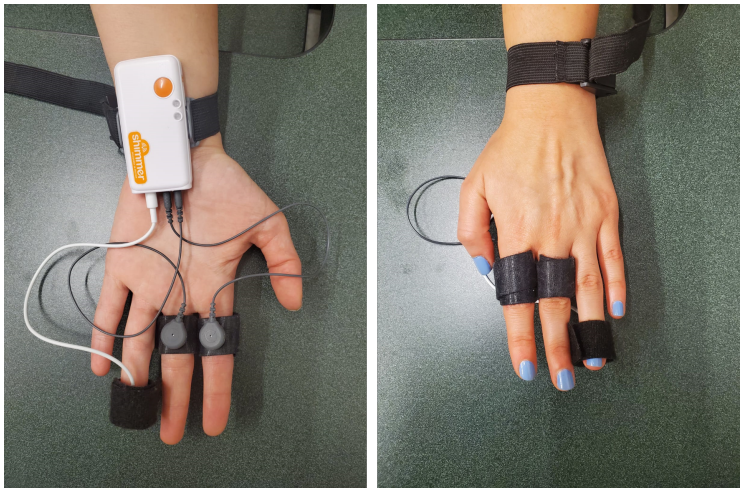


Figure 4.4: Sensor placement of the participant's non-dominant hand from back and front.

naire used the Almere model, developed by Heerink *et al.* [57] as an extended version of the Unified Theory of Acceptance and Use of Technology (UTAUT) [132]. This questionnaire evaluated the acceptance of assistive social technology and comprised ten subscales, including perceived usefulness, sociability, trust, anxiety, and more. The second questionnaire utilized an adapted version of the Technology Familiarity questionnaire [125] to gauge participants' familiarity with robots [51]. This questionnaire focused on measuring participants' familiarity with robots, providing insights into their prior experiences and knowledge regarding robotic technology.

2) *Post-Task Measures:* For the post-task measures, we employed a combination of self-report measures to assess participants' perceptions of the robot used in the study. First, we used the Reliance Intention Scale (RIS) proposed by Lyons *et al.* [87], which measures participants' trust in the robot for future interactions. This scale evaluates participants' inclination to rely on the robot's capabilities and trustworthiness. Second, we utilized the Robotic Social Attributes Scale (ROSAS) developed by Carpinella *et al.* [22]. This scale allows us to gauge participants' perceptions of the robot's social attributes, such as warmth, discomfort, and competence. It provides valuable insights into participants' evaluation of the robot's social presence and qualities.

To evaluate the overall user experience with the robot, we employed the User Experience Questionnaire (UEQ) [72]. The UEQ assesses dimensions such as attractiveness, efficiency, novelty, and dependability, providing a comprehensive understanding of participants' subjective experience. Lastly, we employed the Enjoy Scale (ENJOY) [31] to measure the extent to which participants enjoyed interacting with the robot. This scale

captures participants' level of enjoyment, satisfaction, and positive emotional experience during the interaction.

In addition to these measures, we included specific questions tailored to the experimental context. These questions increased in specificity and depth as participants progressed through the questionnaire. Initially, we asked open-ended and general questions about participants' overall experience, including what they enjoyed and disliked, as well as any specific behaviours exhibited by the robot. We then probed for more specific information related to each norm violation, such as attention checks (participants' awareness of the behaviour), their rating of the severity of the violation, whether the behaviour bothered them, and if it influenced their perception of the robot. Furthermore, we provided an open-ended question for each norm violation to allow participants to elaborate further on their feelings and observations. We applied a thematic analysis method for the open-ended questions, with a single researcher responsible for reading and categorizing the responses to identify recurring themes and patterns. A more detailed description of the analysis method can be found in Section 5.2.1.

3) *Behavioural Measures:* For the collection of behavioural measures, we conducted video analysis to track and observe various behaviours exhibited by the participants. The following behavioural observations were recorded:

i Verbal Response: We quantified the number of audible responses made by the participants throughout the interaction. This measure provides insights into participants' level of engagement and verbal communication during the task.

ii Emotional Response: We categorized and recorded specific emotional responses displayed by the participants. Similarly to Yasuda *et al.* [143], these included anger, surprise, amusement, as well as instances where participants exhibited a neutral or no response. Tracking emotional responses allows us to understand participants' affective states and their emotional reactions to different aspects of the task.

iii Physical Response: We monitored and recorded specific physical responses exhibited by the participants. This included behaviours such as standing up, shifting torso, changes in gaze direction, head gestures, and limb movements. These physical responses provide insights into participants' engagement, attention, nonverbal communication, and affective state during the interaction.

4) *Physiological Measures:* For the collection of physiological data, we utilized the Shimmer3 device to capture Photoplethysmography (PPG) and Galvanic Skin Response (GSR) data. PPG data will be used to extract Heart Rate Variability (HRV) by analyzing

the Root Mean Square of Successive Differences (RMSSD) values, which provide insights into participants' arousal levels (*e.g.,* agitation, stress). Additionally, we extracted Skin Conductance Response (SCR) and Skin Conductance Level (SCL) from GSR data to measure event-driven changes. This was done by analyzing the number of SCR peaks occurring over a given interval (*i.e.,* 10 seconds).

We collected seven events for both physiological sensors in each group. These included a Low Arousal State (LAS) and a High Arousal State (HAS), measured while participants were reading a book and playing the high-paced game 'Fruit Ninja' [52], respectively. The other five events captured Fetch's actions: a non-violating turn, a turn where Fetch rammed into boxes (*box violation*), a turn where Fetch exceeded its time limit (*time violation*), a turn where Fetch made a loud noise upon object retrieval (*noise violation*), and a turn where Fetch stole from the participant's bin (*stealing violation*). For the control group, the same time series was captured during Fetch's turn for comparison.

Chapter 5

Results of User Study

5.1 Participant Perception and Severity Assessment of Norm Violations

Preliminary analyses were conducted using two self-report questionnaires to ensure that the experimental conditions were balanced with respect to participants' familiarity with robots and their perceptions towards assistive social technology. The results from the two-tailed T-Tests did not show any statistically significant differences ($p_n >> 0.05$) between both groups. The extended UTAUT questionnaire, which contains 10 subscales, was analyzed individually and none of which showed a statistically significant difference between the two groups. Furthermore, all subscales of the questionnaire exhibited a normal distribution for the participants, except for "Attitude Towards Technology" (ATT) and "Perceived Enjoyment" (PENJ), which were skewed to the right. This indicates a high attitude towards technology and a high perception of enjoyment among the participants.

As discussed in Measures (4.3), participants were asked to report their perceptions of the norm violations. The results can be seen in Fig. 5.1. As the figure shows, the *box* and *stealing* violations were perceived by 100% of the participants, while the *time* and *noise* violations were perceived by 93% of participants, i.e. two participants did not notice this behaviour when it occurred. An interesting note is that it was not the same participants who did not perceive both violations; it was 4 different participants who did not perceive one violation.

Review of the self-reported severity of the norm violations shows that all norm violations had an average rating greater than 4 on a 7-point Likert scale (with 1 indicating "Not at all"

Table 5.1: Pre-questionnaire analyses

Construct*		**Control**	**Experimental**	**P-Value**
Ext. UTAUT	ANX[1] [4-20]	10.90 ± 3.23	10.93 ± 3.90	0.971
	ATT[2] [3-15]	13.10 ± 2.36	12.34 ± 2.36	0.220
	FC[3] [2-10]	6.23 ± 2.04	6.25 ± 2.19	0.952
	PAD[4] [3-15]	10.58 ± 1.65	10.32 ± 1.81	0.560
	PENJ[5] [5-25]	21.32 ± 3.50	20.35 ± 3.27	0.265
	PEOU[6] [5-25]	13.26 ± 2.29	13.77 ± 2.43	0.393
	PS[7] [4-20]	12.06 ± 2.83	12.19 ± 2.71	0.855
	PU[8] [3-15]	12.19 ± 2.19	11.81 ± 2.30	0.501
	SP[9] [5-25]	10.55 ± 2.99	11.19 ± 3.99	0.474
	Trust [2-10]	6.03 ± 1.74	6.03 ± 1.70	1
Robot Familiarity [9-45]		9.29 ± 2.76	8.19 ± 2.95	0.136

*Subdomains in extended UTAUT: [1]Anxiety, [2]Attitude Towards Technology, [3]Facilitating Conditions, [4]Perceived Adaptiveness, [5]Perceived Enjoyment, [6]Perceived Ease of Use, [7]Perceived Sociability, [8]Perceived Usefulness, [9]Social Presence,

and 7 indicating "A lot"). To determine which norm violations were considered "severe", a one-tailed T-test was conducted with the assumption of a population mean of 4, assuming if an action is deemed norm-violating the average severity would be rated as a 4. The T-test showed significant results for the *noise* and *stealing* violations, with p-values of 0.04 and 6×10^{-6} respectively, indicating that they were perceived as severe. Although both the *box* and *time* violations had an average above 4, the results from the T-test were not significant.

In terms of the "Cognitive Appraisal" (*i.e.,* how it influenced the participant's perception of the robot) and "Disconcertment" (*i.e.,* the level of discomfort or agitation experienced) aspects of the self-reported questions, the participants had a mean equal to or greater than 4 for all norm-violations in terms of *cognitive appraisal*. However, for *disconcertment*, the participants had a mean greater than 4 for only norm violations that directly impacted them (*i.e.,* Time and Stealing).

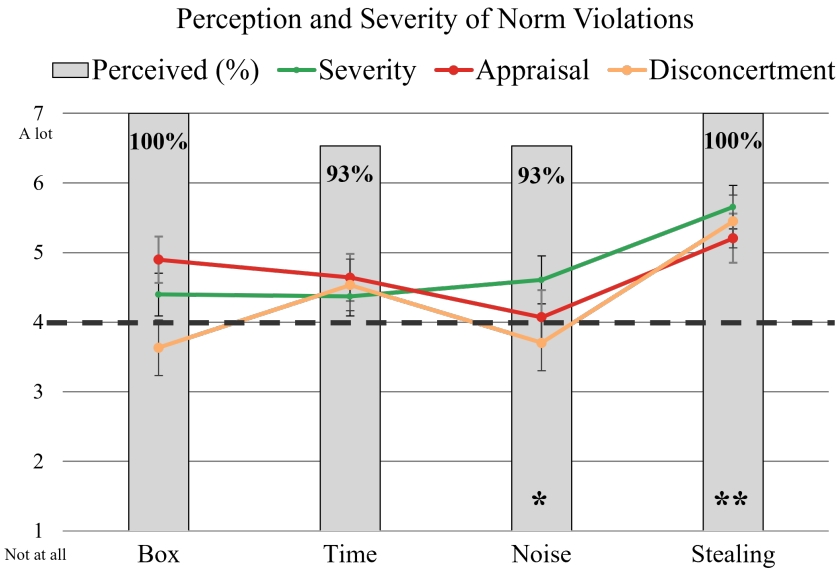


Figure 5.1: The figure presents the percentage of norm violations perceived by participants, the severity ratings, the impact on participants' perceptions of the robot (*i.e.,* cognitive appraisal), and the degree to which it discomforted or agitated them (*i.e., disconcertment*). Norm violations rated as severe (*i.e.,* significantly greater than 4) with a significance level of $p < 0.05$ or $p < 0.01$ are marked with * and **, respectively.

5.2 Perceptual, behavioural, and Physiological Responses to Norm Violations

5.2.1 Trust, Discomfort, Competence, and Enjoyment

The effects on trust were evaluated using the RIS questionnaire which showed significantly lower trust in the experimental group after the experiment. Results from a two-tailed T-Test, as presented in Table 5.2, indicated a significant difference between the experimental and control groups with a p-value of 0.046. It is noteworthy that the extended UTAUT pre-questionnair's Trust subscale revealed that both groups had similar mean scores prior to the experiment. The overall trust and the difference between the two groups can be visualized in Figure 5.2.

The ROSAS questionnaire contains subscales of Discomfort, Competence, and Warmth, in which a two-tailed T-Test was conducted for all three. The results in Table 5.2, and illustrated in Fig. 5.3, show the Discomfort scale was the only one with a significant dif-

Table 5.2: Results from post-questionnaires measuring trust, social attributes, user experience, and enjoyment towards the robot in the experimental and control groups. A significant difference, $p < 0.05$ and $p < 0.01$, are marked by * and ** respectively. Table from [75]

Construct		**Control**	**Experimental**	**P-Value**
Reliance Intention Scale (RIS) [5-25]		14.65 ± 4.22	12.57 ± 3.70	***0.046**
ROSAS	Discomfort [6-30]	11.55 ± 4.13	14.80 ± 3.89	****0.0025**
	Warmth [5-25]	13.23 ± 3.21	13.3 ± 3.62	0.933
	Competence [5-25]	18.94 ± 3.19	17.87 ± 3.00	0.183
User Experience (UEQ) [8-40]		30.45 ± 3.77	28.97 ± 5.02	0.195
The Enjoy Scale (ENJOY) [9-45]		37.58 ± 4.63	35.03 ± 4.94	***0.042**

ference between the groups with a p-value of 0.0025. While Competence trended in the expected direction of *H2b*, the difference was not significant. Warmth did not show any changes between the two groups.

The UEQ questionnaire was used to measure the comprehensive impression of the user experience, and the ENJOY questionnaire was used to specifically distinguish a difference in the overall enjoyment between the two groups. The results from the UEQ questionnaire did not show a significant difference. Conversely, the results from the ENJOY questionnaire showed a significant difference between the two groups, with a p-value of 0.042, indicating the experimental group experienced a significantly lower level of enjoyment. This is illustrated in Fig. 5.2 where we see that both groups rated their enjoyment relatively high, however, the experimental group reported significantly lower levels.

Open-Ended Questions

The analysis of the open-ended questions used a qualitative approach to gain deeper insights into participants' perceptions and reactions to the norm violations. We applied a thematic analysis method where a single researcher read and categorized the responses to identify recurring themes and patterns. The analysis aimed to capture salient expressions conveyed by participants in response to the norm violations, including a range of emotions and impactful reactions. Keywords and phrases related to shock and emotional reactions (*e.g.,* "surprised", "frustrated", "angry", etc.) were identified and used as categories for organizing the data. With the data categorized within the norm violations, the spectrum of respective emotional responses and the severity could be analyzed.

The analysis revealed a distinct difference between the experiment and control groups, largely influenced by the robot's social norm violations. Participants expressed surprise,

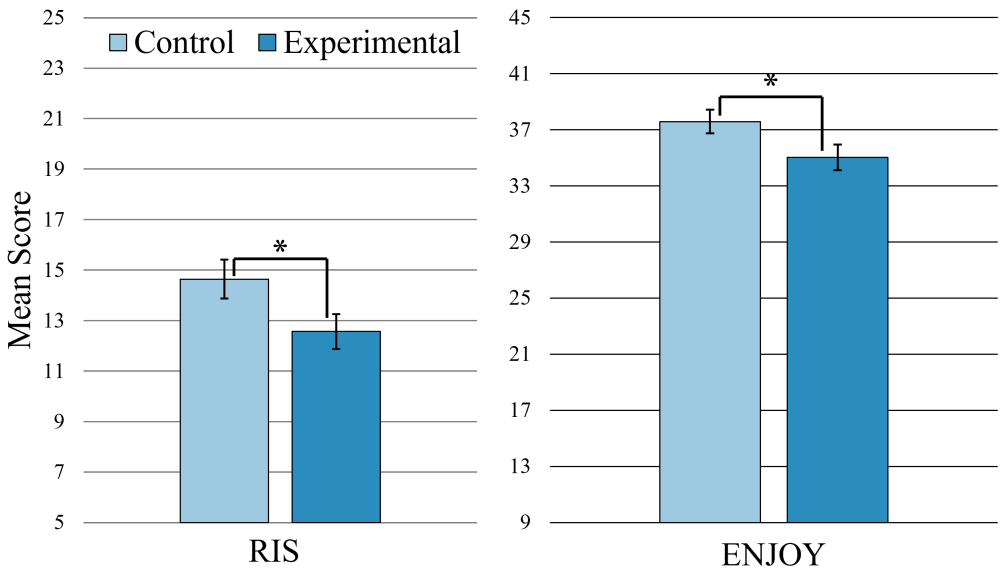


Figure 5.2: The outcomes of both the RIS and ENJOY questionnaires, displaying the relative scales for each on the y-axis. Consequently, the y-axis minimum represents the lowest possible score for each questionnaire, while the maximum corresponds to the highest attainable score. Thus, greater values on the y-axis signify increased levels of trust and enjoyment. The results show that the experimental group had significantly lower mean trust and enjoyment scores compared to the control group. A statistically significant difference of $p < 0.05$ is marked by *.

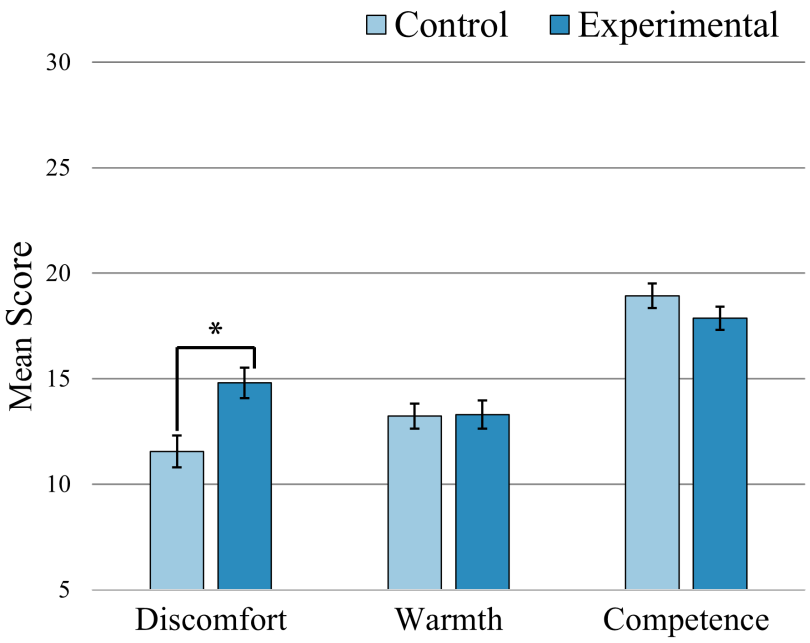


Figure 5.3: The mean scores for the RoSAS questionnaire's subscales depict the participants' relative discomfort, warmth, and competence towards the robot. The y-axis represents the range of possible scores achievable using a 5-point Likert scale. A significant difference of $p < 0.05$ is marked by *.

shock, and other negative emotional reactions to Fetch's norm violations. For example, many participants stated they experienced shock and surprise during the *box* violation when Fetch collided with the fragile boxes. Similarly, participants expressed frustration, impatience, and a sense of unfairness when Fetch exceeded its allotted time to find an object during the *time* violation.

Fetch's *noise* violation caused mixed responses, ranging from amusement to bewilderment, highlighting the effects of unexpected robot behaviour. The *stealing* violation, where Fetch took an object from a participant's collection, was described as "stealing" by participants and triggered strong negative reactions. Some reported annoyance and a sense of betrayal, while others found it humorous, showcasing the diverse range of human responses to robot norm violations in this experiment.

In contrast, the control group's responses, who did not encounter any norm-violating behaviours, focused more on the robot's operational aspects. These included object retrieval and speed, with a desire for increased interaction and communication.

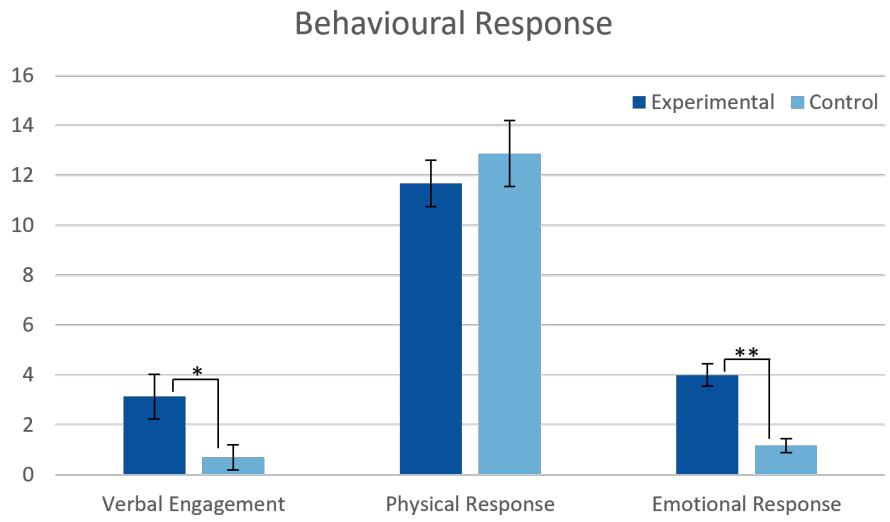


Figure 5.4: Differences in behavioural observations captured during the analysis of the experimental and control groups. The three categories of behavioural responses—verbal, physical, and emotional—are shown. The bars represent the mean frequency of each type of response, with error bars indicating the standard error of the mean (SEM). Significance levels of $p < 0.05$ or $p < 0.01$ are marked with * and **, respectively.

5.2.2 behavioural Responses

We conducted a behavioural analysis by reviewing recorded videos of participants from the control and experimental groups. Specifically, we focused on the period when the robot searched for its item during the game. Three independent video coders assessed the recorded videos for the video coding process. The video coders underwent training collectively, which involved watching videos from three participants and coding them using a spreadsheet rubric. After completing the video, the coders compared their coding answers and discussed any discrepancies. These discrepancies helped refine and solidify the definition of the behaviours being identified. Through discussions, the coders clarified their understanding of the coding criteria and reached a consensus on accurately classifying and recording the observed behaviours.

To assess the agreement among the video coders, we calculated inter-rater reliability scores on 15% of the observed videos. The results, measured using Fleiss Kappa [41] ($\kappa = 0.64$) and Krippendorff's alpha [71] ($\alpha = 0.64$), indicated a satisfactory level of agreement among the coders.

Our behavioural analysis followed a schema that included three categories: Verbal En-

gagement, Emotional Response, and Physical Response. Verbal Engagement was quantified by the total number of audible responses made by the participants, while Emotional Response was assessed by counting emotional facial expressions beyond a neutral expression, including anger, surprise, and amusement. Physical Response involves tracking activities such as standing up, shifting the torso, gaze change, head gestures, and limb movements. To track these responses, we used a spreadsheet and our coding scheme to record the occurrences of each category.

T-Tests revealed significant differences in Verbal Engagement ($p = 0.03$) and Emotional Response ($p = 2.8^{-10}$) between the control and experimental groups. These findings suggest that participants in the experimental group exhibited higher levels of engagement and displayed more emotional responses. However, the two groups had no statistically significant difference in Physical Responses. Refer to Fig. 5.4 for a visual representation of these results.

Further analysis of the Emotional Response category looking at the subcategories anger, surprise, and amusement, revealed statistically significant differences between the control and experimental groups. The respective p-values of 0.02, 4.2^{-6}, and 4.0^{-5} indicate a statistically higher occurrence of all three emotional reactions in the experimental group.

In summary, our behavioural analysis highlighted notable differences between the control and experimental groups, particularly in terms of verbal engagement and emotional responses. In particular, the findings suggest that the "cheating" behaviour programmed into the robot for the experimental group elicited more intense reactions from the participants.

5.2.3 Physiological Responses

While we conducted analysis for 60 participants for the questionnaire and behavioural analysis, we were unable to conduct physiological analysis on the entire sample. Data from the initial 20 participants was unusable due to a faulty PPG sensor. Additional discards related to technical issues, motion artifacts, and sensor placement instability resulted in a total of 25 participants (13 experimental, 12 control) whose data was included in our analysis.

In our physiological data collection, we measured PPG and GSR during Fetch's turns in both groups. This was done to monitor any changes in physiological states that may have resulted from the norm violations and compare them to those from the control group over the same time series.

PPG allowed us to measure Heart Rate Variability (HRV) of the participants, using changes in the Root Mean Square of Successive Differences (RMSSD) as an indicator of arousal, *e.g.,* agitation and/or stress. A lower RMSSD value signifies a higher state of arousal. With GSR, arousal was determined by the number of Skin Conductance Response (SCR) peaks per 10 seconds, with more peaks indicating increased arousal. Consequently, the two metrics, RMSSD and SCR, are anti-correlated but jointly provide a comprehensive measure of an individual's physiological state changes.

We gathered data from seven distinct events using two physiological sensors in each group. This included measuring Low and High Arousal States during passive reading and the high-paced game 'Fruit Ninja' [52], respectively. The other five events documented Fetch's actions, including non-violating turns and four separate norm-violating turns (*i.e., box violation*, *time violation*, *noise violation*, and *stealing violation*) and the corresponding time series in the control group.

Due to the smaller sample size, we conducted a T-Test analysis supported by Wilcoxon analysis. The T-Test and Wilcoxon analyses of HRV data revealed no statistically significant differences between Fetch's turns and the Low Arousal State (LAS) in the control group. However, all but one (*stealing* violation) of Fetch's turns were significantly different from the High Arousal State (HAS) (p¡0.01, after Bonferroni correction). In the experimental group, although not statistically significant, there was a trend of higher arousal (lower RMSSD) during the four norm-violating turns.

Regarding the GSR data analysis, in the experimental group, a higher number of SCR peaks were observed per 10 seconds during norm-violating turns compared to the control group. The difference was statistically significant for all four norm violations when compared to the LAS (*Box*: $p = 9.8^{-5}$, *Time*: $p = 9.0^{-7}$, *Noise*: $p = p = 9.9^{-7}$, *Stealing*: $p = 2.9^{-5}$). When compared to the HAS, there was no significant difference except for the *box* violation, p=0.003. In the control group, no significant difference was observed when compared to the LAS, while all were significantly different from the HAS (*Box*: $p = 4.1^{-5}$, *Time*: $p = 2.0^{-5}$, *Noise*: $p = 2.8^{-7}$, *Stealing*: $p = 9.7^{-7}$).

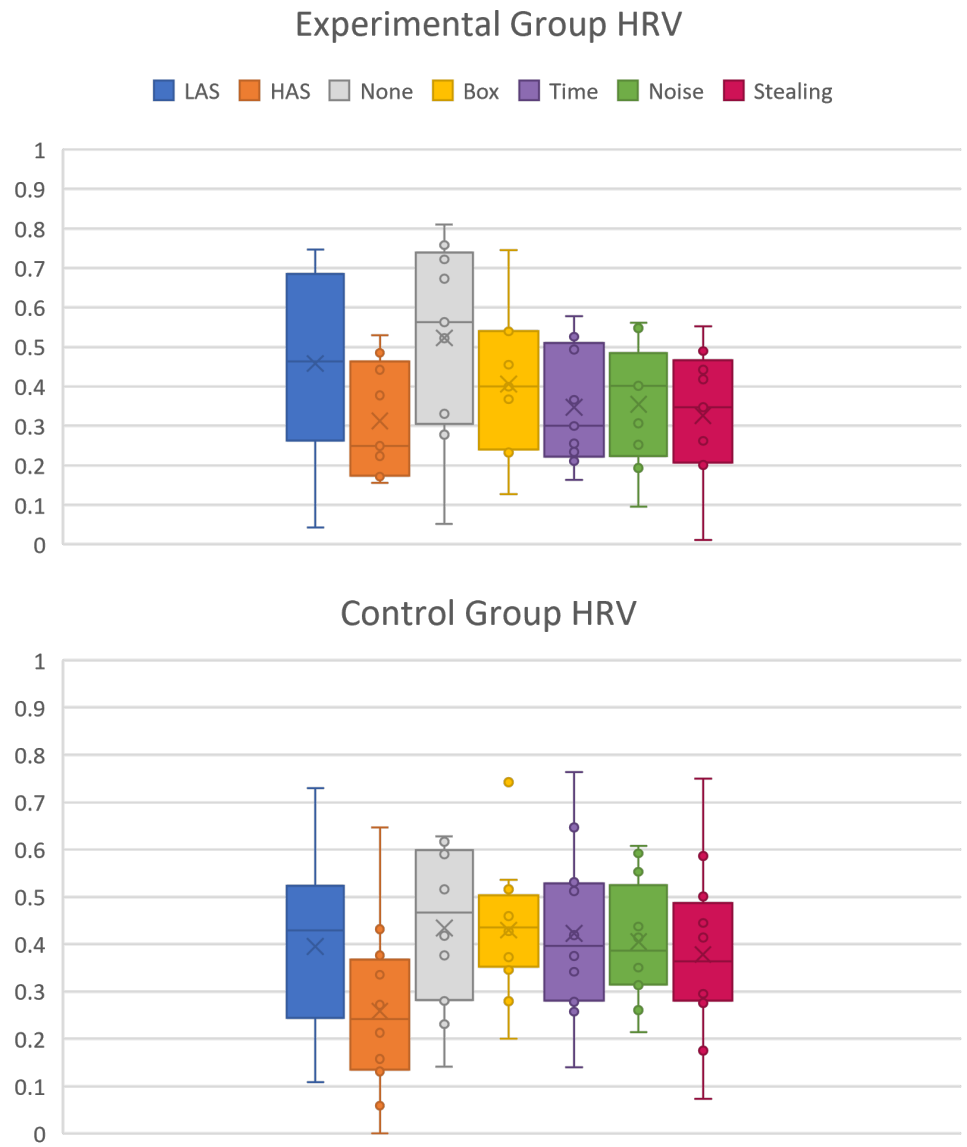


Figure 5.5: Comparison of Heart Rate Variability (HRV) in the two conditions, measured by RMSSD (Root Mean Square of Successive Differences). Comparing the norm violation time series to the Low Arousal State (LAS) and High Arousal State (HAS), we see the RMSSD exhibits a lower tendency (closer to the HAS) in the experimental group compared to the control group indicating a higher state of arousal (*e.g.*, agitation and stress). Whereas the control group's corresponding time series exhibit higher RMSSD values similar to the LAS, indicating a lower state of arousal. The data shown in the graph has been normalized using a min/max normalization method to accurately compare data from different participants.

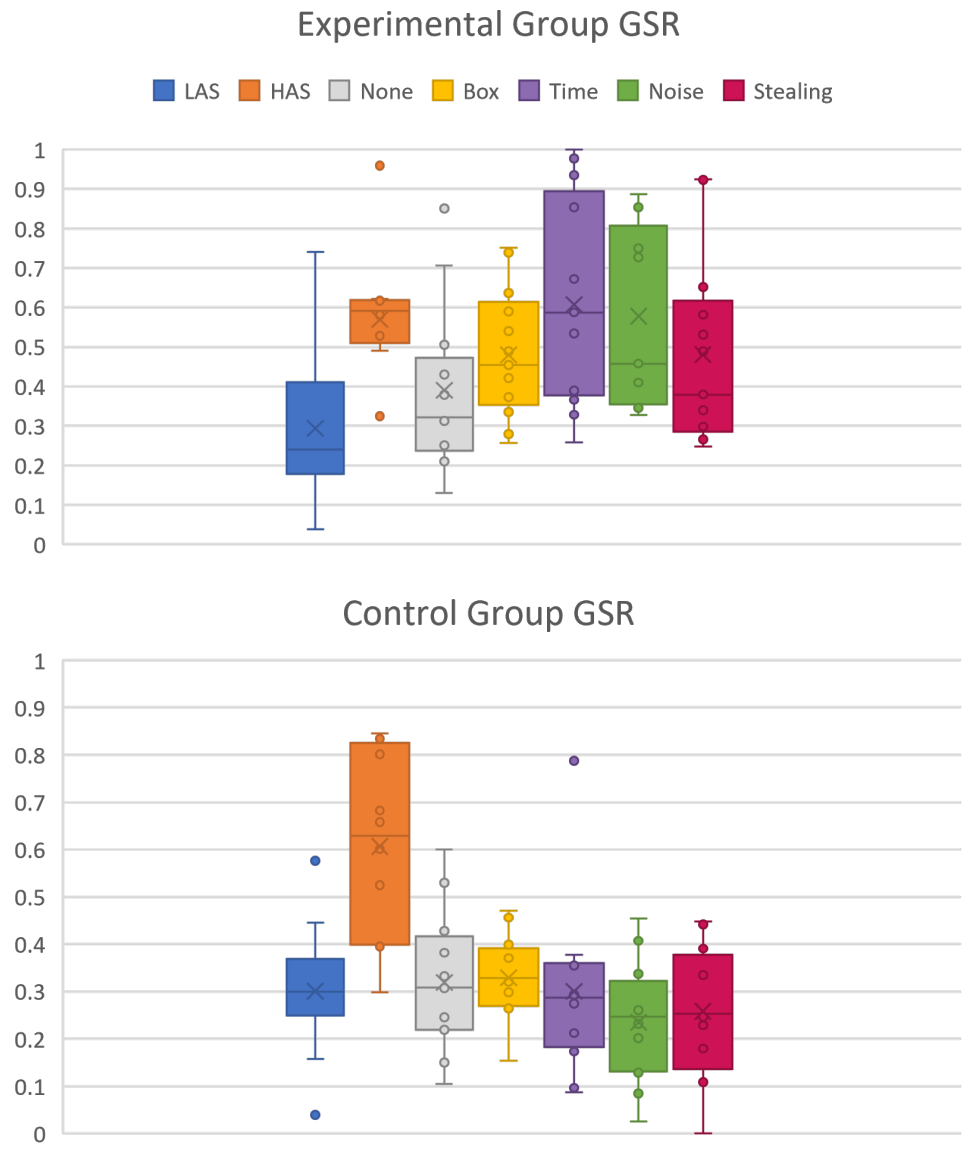


Figure 5.6: Comparison of Galvanic Skin Response (GSR) using Skin Conductance Response (SCR) peaks per 10 seconds between the experimental and control group. Comparing the Low Arousal State (LAS) and High Arousal State (HAS) to the four norm violation time series we see a significantly higher number of SCR peaks in the experimental group compared to the control group. This indicates participants in the experimental group had an SCR peak frequency similar to the HAS during the norm-violating time series, whereas, the participants in the control group had an SCR peak frequency similar to the LAS for the same corresponding time series. This supports the HRV data in Fig. 5.5 that the participants in the experimental conditions are in a higher arousal state. The data shown in the graph has been normalized using a min/max normalization method to compare data from different participants accurately.

Chapter 6

Discussion and Implications

6.1 Preliminary Analysis and Evaluation of Participant Responses

The preliminary analyses suggested the sample group used for this experiment to be homogeneous with normally distributed data with the exceptions of Attitude Towards Technology (ATT) and Perceived Enjoyment (PEN). These two subscales were skewed to the right indicating the majority of participants had a high attitude toward technology and a higher expectation of enjoyment. We hypothesize that this is due to the fact the participants volunteered to take part in the experiment, therefore, they are likely to have a positive attitude towards technology and expected the experience to be enjoyable. Additionally, the recruitment of participants took place at a university, which may have resulted in a population with a higher inclination towards technology, thus leading to a higher perception of enjoyment with respect to an experiment involving robots.

Perception and Severity of the Norm-Violations

To confirm that the study's findings were attributable to the norm-violating behaviour, we posed several manipulation checks to ensure the participants in the experimental group noticed and perceived the behaviour as a social norm violation. Specifically, we asked if they noticed the behaviour, and if so, the perceived severity of the behaviour, its impact on participants' perceptions of the robot, and whether it caused discomfort or annoyance.

From these questions, we found the experimental participants did notice the norm-violating behaviour and rated them high in terms of severity.

An illustration of these findings can be seen in Fig. 5.1. All violations received a mean severity rating above 4 and the severity increased monotonically. This increase may be attributed to a compounding effect from participants experiencing a series of violations, or because the more recent norm violations were fresh in their minds when they completed the questionnaire. In designing the experiment, we assumed that the violations that had a direct impact on the participant, such as the *time* and *stealing* violations, would be ranked the most severe as they have the ability to alter the outcome of the game. However, one reason we think the *time* violation did not support this hypothesis is that at the time of the violation, most participants had already searched the room thoroughly, and they knew the location of the next object displayed on the screen, and thus might have been less concerned when losing a small amount of their allocated turn.

In terms of the *cognitive appraisal* and *disconcertment* aspects of the self-reported questions, we found some expected results. The participants' ratings for disconcertment by the norm violations showed that those that had a direct effect on the participant and the outcome of the game (*i.e., time* and *stealing* violations) were rated higher than the behaviour that did not. As for the effect on participants' cognitive appraisal of the robot, the results suggest that the violation of injunctive norms (*i.e., box* and *stealing* violations) had a greater impact than descriptive norms (*i.e., noise* and *time*). This could be due to injunctive norms being on behaviours that are perceived to be either approved or disapproved of by others and therefore violations of these norms are likely to have a stronger effect on a person's perception.

6.1.1 Questionnaire Analysis: Assessing Trust, Discomfort, Competence, and Overall Enjoyment

The results of the post-questionnaires revealed some insights into the effects of a robot's violation of social norms on human participants. A robot exhibiting norm-violating behaviour seems to lead to lower trust, lower overall enjoyment, and higher discomfort among the participants.

RQ1: How do social norm violations affect trust in robots? The results from the RIS questionnaire showed that trust in the experimental condition was significantly lower than in the control group. This supports our hypothesis, *H1*, that a robot performing social norm violations would lower trust in robots during human-robot interactions. This difference may be attributed to the robot's actions being inconsistent with user expectations

causing them to question the robot's actions and ability to be reliable and trustworthy in future encounters. The violation of injunctive norms may have had a greater impact on trust because these norms are perceived to be approved or disapproved of by others, such as not stealing or not causing damage to others' items. Violating such norms usually carries more severe negative social consequences than descriptive norms. When these injunctive norms are violated, it can lead to a stronger negative emotional response, including feelings of anger or betrayal, which can, in turn, lead to lower trust in the robot.

RQ2: How are discomfort, and perceived competence affected by robots performing social norm violations? The results from the ROSAS questionnaire showed that participants in the experimental condition experienced a significantly higher level of discomfort than the control group. This supports hypothesis *H2a* that participants would experience a higher level of discomfort when interacting with a robot violating social norms. The norm violations used in this experiment may have caused discomfort because they involved behaviours that are typically disapproved and directly impacted the participants' own ability to perform well in the game. For *H2b*, while not statistically significant, the perceived competence of the robot was reported lower in the experimental group. The robot used speech to indicate the intent behind its actions in both groups. This was intended to make the norm-violating behaviour, such as running into the boxes, appear deliberate to elicit a norm-violation response from the participants instead of being viewed as an error due to incompetence. This may be why we do not see a significant difference in competence. Additionally, the norm-violating actions resulted in the robot obtaining the object and winning the game which may have also affected its perceived competence.

RQ3: How is overall enjoyment affected when robots perform social norm violations? The results from the ENJOY scale showed a significant difference between the experimental and control group. Participants in the experimental group reported a significantly lower level of enjoyment than the control group. This finding is consistent with our hypothesis, *H3*, that social norm violations during human-robot interactions can have a negative impact on enjoyment. When a robot performs actions that are perceived as socially inappropriate or disrespectful, it may trigger negative emotional responses such as discomfort, unease, or even frustration. Such emotional responses can then influence the overall enjoyment of the interaction, leading to lower ratings on measures such as the ENJOY scale. Additionally, violating social norms can disrupt social expectations and trust that are crucial to creating a positive user experience.

Qualitative Insights: Participants' Perspectives from Open-ended Questions

Answers to open-ended questions further highlighted the negative effect of the robot's norm violations. Participants expressed surprise, shock, and different negative emotions to the norm violations, supporting **H2a** (increase in discomfort) and **H3** (reduction in overall enjoyment). For example, one participant in the experimental group commented on the *stealing* violation, stating, "I felt very annoyed because he took my cup, made fun of me, and then took away my points and said he won the game". This comment shows the effect of unexpected behaviour and norm violations on participants' reactions, especially when the violation directly affected them. Another participant shared their reaction to a norm violation that did not directly affect them, saying, "I was shocked because it was clearly mentioned on the box not to move it. The way the robot knocked the box off was funny". These varied responses suggest that individual characteristics and situational factors affect the perception of social norm violations, adding complexity to these interactions. The analysis of the open-ended questions showed a variety of emotional reactions, all centred around surprise or shock. These results emphasize the need to consider individual responses when designing and programming robots to work within social boundaries. Unexpected deviations from norms can cause a wide range of reactions, but they consistently cause shock, underlining the importance of following social norms to promote positive human-robot interactions.

6.1.2 Examining Behavioural Responses

Our behavioural analysis was a key part of the study, providing additional data that added to the perspectives we gained from the self-reported questionnaires, open-ended questions, and physiological analysis. This additional layer of observational data provided another way to address our research questions and support our hypotheses. To draw conclusions, we considered the three modes of behavioural responses: Verbal, Emotional, and Physical. Regarding RQ1, the increased verbal engagement and emotional responses observed in the experimental group could be interpreted as reactions to reduced trust in the robot because of the norm violations. These results back up the findings from the self-reported trust measures, providing extra evidence that participants' trust in the robot was negatively affected by its norm violations supporting **H1**. Regarding RQ2, the range of emotional reactions we observed and coded further supports **H2a** related to discomfort. The fact that participants in the experimental group showed a wider variety of emotional responses could suggest increased discomfort due to the norm violations.

Regarding RQ3, the stronger emotional responses could also suggest reduced overall enjoyment, supporting **H3**, due to the unexpected and disruptive behaviour of the robot. The surprise or frustration shown by the participants in the experimental condition (with robot norm violations) provides observational support to the notion that social norm violations by robots can decrease the enjoyment of human-robot interactions. Finally, while the behavioural analysis did not directly measure physiological changes (RQ4), the observation of stronger emotional responses in the experimental group suggests possible agitation or stress, indirectly supporting **H4** that norm violations could cause physiological state changes. However, this suggestion needs further investigation through direct physiological measurements to draw definitive conclusions. This mixed-method approach highlights the importance of using different response evaluation methods in HRI research. The behavioural analysis provided valuable insights, especially with regard to verbal and emotional responses, showing that the dynamics of HRIs are complex and should be examined in multiple ways.

6.1.3 Examining Physiological State Changes

The physiological analysis provided a perspective on the more subconscious reactions of participants, allowing us to observe the detailed relationship between robot social norm violations and participants' physiological state changes (RQ4). The physiological measures specifically focused on changes in the arousal level, which could reflect participants' agitation or stress [64, 6] caused by the robot's norm violating behaviour. Our combined HRV and GSR data analysis gave us valuable insights into the subtle influence of social norm violations on the participants' arousal levels.

The control group's HRV data revealed no significant differences between the Low Arousal State (LAS) and the four event time series, representing the time periods when norm violations would have occurred in the experimental group. This suggests that participants maintained a similar level of low arousal during the robot's turn as they did during the LAS. When we compared these same four time series to the control group's High Arousal State (HAS), three of the four events (*i.e., box*, *time*, and *noise* violations) showed a significant difference in terms of RMSSD (p¡¡0.01). The fourth event, *stealing*, had a small p-value, suggesting a trend in RMSSD in the expected direction.

These findings suggest that the control group's arousal states during the robot's turn were very similar to the LAS. The experimental group, however, showed the opposite trend in HRV, as anticipated. Although the analysis did not yield a statistically significant difference between the norm violation events and the LAS, the p-values were low. This

suggests a trend towards a higher arousal state during these events among participants in the experimental condition.

The comparison between the HAS and the norm violation events yielded no statistically significant differences. This suggests that the Root Mean Square of Successive Differences (RMSSD) values were similar to the HAS during these events. These results support our hypothesis (H4), namely that robot social norm violations can lead to physiological changes indicative of increased arousal. They also corroborate our second hypothesis (H2a), suggesting that norm violations can cause discomfort, as the observed arousal could be interpreted as agitation or stress. Moreover, it shows that even subtle, unexpected behaviours from social robots can influence human users on a physiological level, adding another dimension to the user experience in human-robot interactions.

GSR data reinforced these findings, showing a higher frequency of SCR peaks during instances of norm-violating robot actions within the experimental group. Notably, the GSR data for the experimental group demonstrated a significant difference between the LAS and all norm violation events, pointing to increased arousal during these periods.

Additionally, there was a notable difference in the Skin Conductance Response (SCR) peak frequency during the robot's first turn, where no violation occurred, compared to the High Arousal State (HAS). This difference strengthens the argument that this increase in SCR peak frequency, and thus arousal, resulted from the norm violations, rather than from the game itself.

Overall, these findings suggest an increased arousal state among participants in the experimental condition during episodes of norm violation, echoing our observations from the HRV data. The GSR results from the control group further highlight these assertions as they did not display this increased SCR peak frequency during the same time periods.

The analysis found no statistically significant differences between the LAS and all four time periods within the control group. However, when these periods were compared to the HAS, all four times showed a significant difference (p¡¡0.01). These complementary findings suggest that the participants in the control group did not reach a high arousal state during any of the robot's turns, underscoring the impact of norm-violating behaviour on arousal levels.

To summarise, the two physiological measures indicate that participants in the experimental group experience a higher state of arousal, therefore they underwent physiological state changes which supports our hypothesis (**H4**). Additionally, these physiological state changes may indicate a higher level of stress/agitation which complements our questionnaires results in supporting H3.

6.2 Interpretation of Findings and Future Directions

Our study used a mixed methods approach, including questionnaires, open-ended questions, behavioural and physiological analyses, to explore the complex effects of social norm violations in HRI. This mixed-method approach, providing quantitative and qualitative data, allowed for a thorough evaluation of our hypotheses and exploration of our research questions. In our questionnaire responses, we found that robot norm violations significantly lowered participant trust (H1), increased discomfort (H2a), and reduced overall enjoyment (H3). These findings were consistent, confirming the negative impact of such behaviours on the user experience and highlighting the essential role of social norms in HRI.

Supplementing the questionnaire data, the open-ended questions provided nuanced insights into participants' subjective experiences. The participants reported 'surprise' and 'shock' in response to the robot's norm violations further emphasizing the intensity of these negative impacts. This pointed towards individual variations in responses to norm violations, affirming that while everyone may not respond similarly, norm violations universally cause strong reactions. The open-ended questions also showed that even indirect violations (*e.g.,* box violation) could still lead to strong emotional reactions.

The behavioural analysis gave additional evidence for our hypotheses. Increased verbal engagement and emotional responses in the experimental group supported the idea of decreased trust (H1), increased discomfort (H2a), and reduced overall enjoyment (H3). While the physical response did not show any significant difference between the groups, the difference between emotional and physical reactions emphasized the importance of evaluating different behavioural response types in HRI.

Finally, the physiological analysis provided an objective measure to examine the subconscious, physical effects of the norm violations. The results revealed significant physiological state changes resulting from the norm violations, thereby supporting **H4** (*i.e.,* norm violations will elicit physiological state changes), and highlighting the subconscious impacts that these violations can have on individuals.

By combining these complementary methods of analysis using a mixed methods approach, our study strongly supported our hypotheses, namely affirming the significant impact of social norm violations in HRI. The varied emotional, verbal, and physiological reactions obtained in our specific study indicate that norm violations can negatively affect users' experience and perception of social robots. This underscores the need for further research into social norms in HRI, with a view toward better programming and design of social robots that can respect these norms and encourage more positive, engaging interactions.

The findings of our study open up several future directions in the field of HRI. One such potential investigation lies in examining the role of the severity of social norm violations on human responses. Our current work evaluated various types of norm violations but did not expressly focus on their severity. Future research could systematically manipulate the severity of violations, from minor to severe, to understand its impact on trust, discomfort, enjoyment, and physiological arousal. This could help to create a more comprehensive picture of how humans perceive and react to different degrees of norm-violating behaviour by robots. Another promising area of investigation involves exploring the long-term effects of repeated exposure to norm violations by robots. Would individuals become used to such behaviour over time, thereby leading to diminished emotional and physiological responses? Or would the negative effects persist, or even get worse, with repeated exposure to the violations? Understanding these dynamics could have big implications for the design and programming of social robots, particularly those intended for long-term or repeated interactions with humans. Furthermore, we could delve deeper into the individual differences in reactions to norm violations. Our open-ended responses suggested varied emotional reactions, indicating that personal characteristics and situational factors significantly influence how individuals perceive and respond to norm violations. Future research could focus on identifying these factors, potentially using personality traits or past experiences as predictors for reactions to norm violations. Thus, while our study provides insights into the effects of social norm violations in HRI, it also opens up a wide range of exciting future research directions, promising to deepen our understanding of these complex interactions. Such research could further our ability to design and build social robots that respect human norms and values, thereby enhancing the HRI experience.

Chapter 7

Limitations and Conclusion

While this study provides valuable insights into the effects of social norm violations in human-robot interactions, several limitations must be acknowledged.

Firstly, our sample size was limited to 60 participants, split equally between the experimental and control groups. While this number allowed for statistical analyses and the observation of significant effects, a larger sample size would provide a more robust examination of the effects and increase the generalizability of our findings. However, running an in-person HRI study with a complex scenario involving a mobile manipulator robot, cannot be done easily with hundreds of participants in the typical time frame of an HRI project, let alone running such a study in a naturalistic application scenario (e.g. in a home or industrial setting). Additionally, the sample size for the physiological analysis was reduced further due to an improper PPG sensor, technical issues, motion artifacts, and sensor placement instability. The smaller sample size has limited our statistical significance.

Secondly, our study employed a specific robot, Fetch, with unique design features, mobility, and dexterity, potentially limiting the applicability of results to other robots with distinct designs and functionalities. Moreover, we utilized a defined set of norm-violating behaviours, which might not represent all possible violations in various HRI scenarios.

Thirdly, the study context, a laboratory-based competitive game, might have influenced the intensity of participants' emotional and physiological responses. Hence, our findings may not be directly applicable to less competitive, more familiar, or day-to-day HRI settings.

Next, the study was designed as a between-participants experiment, with participants randomly assigned to control and experimental groups. Although this design aimed to

control for potential confounding variables, the influence of unmeasured factors such as participants' social or cultural backgrounds on their perceptions and responses could not be completely ruled out.

Moreover, our study did not explore the impacts of repeated or prolonged exposure to norm violations or varying degrees of severity in these violations, leaving room for further investigation into these dimensions of HRI.

Lastly, the study employed a combination of Wizard of Oz control and autonomous features, while giving an impression of full autonomy to the participants. The understanding of human involvement in robot operations may have influenced participants' responses and warrants further exploration.

Future research should aim to address these limitations by exploring a wider range of norm violations, employing larger and more diverse participant samples, investigating the impacts of repeated exposure and varying severity of norm violations, and being more transparent about the extent of human control in robot operation.

Conclusion

Our investigation into the effects of social norm violations by robots on human perceptions and attitudes offers valuable insights into the field of Human-Robot Interaction (HRI). Utilizing the Fetch robot in a competitive scavenger hunt game scenario, we programmed specific behaviours representing injunctive and descriptive norm violations. Employing a 2x1 between-participants experimental design, participants were randomly assigned to an experimental group, exposed to social norm violations, or a control group devoid of such violations. Our mixed-method approach, combining questionnaires, behavioural observations, and physiological measurements, revealed significant findings.

Participants in the experimental group reported a notable decrease in trust and overall enjoyment along with an increase in discomfort, underscoring the negative implications of a robot's norm-violating behaviour. Our findings resonate with the importance of considering social norms during the design and programming of robots to enhance user experiences. Furthermore, our results emphasize that predicting and understanding how norm violations impact users' trust, comfort, and enjoyment can assist designers and programmers in creating robots that respect human norms, fostering a more productive interaction.

In conclusion, this study contributes valuable insights to the expanding dialogue on the role and impact of social norm violations in human-robot interactions. As the integration of social robots into various aspects of human life advances, the exploration of these

nuanced interactions remains increasingly critical. By elucidating how norm violations affect human perceptions and attitudes, we can contribute to the development of robots that can integrate more seamlessly into human society, promoting positive human-robot interactions.

References

[1] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[2] Chatchalita Asavanant and Hiroyuki Umemuro. Personal space violation by a robot: An application of expectation violation theory in human-robot interaction. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1181–1188. IEEE, 2021.

[3] João Avelino, Leonel Garcia-Marques, Rodrigo Ventura, and Alexandre Bernardino. Break the ice: a survey on socially aware engagement for human–robot first encounters. *International Journal of Social Robotics*, 13(8):1851–1877, 2021.

[4] Ali Ayub, Huiqing Hu, Guangwei Zhou, Carter Fendley, Crystal Ramsay, Kathy Lou Jackson, and Alan R Wagner. If you cheat, i cheat: Cheating on a collaborative task with a social robot. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 229–235. IEEE, 2021.

[5] Negin Azizi, Kevin Fan, Melanie Jouaiti, and Kerstin Dautenhahn. Play dynamics in a collaborative game with a robot as a play-mediator. In *International Conference on Social Robotics*, pages 415–426. Springer, 2022.

[6] Jorn Bakker, Mykola Pechenizkiy, and Natalia Sidorova. What's your current stress level? detection of stress patterns from GSR sensor data. In *IEEE ICDM*, pages 573–580. IEEE, 2011.

[7] Adrian Ball, David Rye, David Silvera-Tawil, and Mari Velonaki. Group vs. individual comfort when a robot approaches. In *International Conference on Social Robotics*, pages 41–50. Springer, 2015.

[8] Santosh Balajee Banisetty. *Multi-context socially-aware navigation using non-linear optimization*. PhD thesis, University of Nevada, Reno, 2020.

[9] Santosh Balajee Banisetty and Tom Williams. Implicit communication through social distancing: Can social navigation communicate social norms? In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 499–504, 2021.

[10] Suna Bensch, Aleksandar Jevtic, and Thomas Hellström. On interaction quality in human-robot interaction. In *ICAART 2017 Proceedings of the 9th International Conference on Agents and Artificial Intelligence, vol. 1*, pages 182–189. SciTePress, 2017.

[11] Mriganka Biswas and John Murray. The effects of cognitive biases in long-term human-robot interactions: case studies using three cognitive biases on marc the humanoid robot. In *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings 8*, pages 148–158. Springer, 2016.

[12] Guido Boella and Rossana Damiano. An architecture for normative reactive agents. In *Intelligent Agents and Multi-Agent Systems: 5th Pacific Rim International Workshop on Multi-Agents, PRIMA 2002 Tokyo, Japan, August 18–19, 2002 Proceedings 5*, pages 1–17. Springer, 2002.

[13] Guido Boella, Leendert Van Der Torre, and Harko Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17:1–10, 2008.

[14] Annika Boos, Markus Zimmermann, Monika Zych, and Klaus Bengler. Polite and unambiguous requests facilitate willingness to help an autonomous delivery robot and favourable social attributions. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1620–1626. IEEE, 2022.

[15] Cynthia Breazeal. *Designing sociable robots*. MIT press, 2004.

[16] Cynthia Breazeal. Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3527–3538, 2009.

[17] Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. *Springer handbook of robotics*, pages 1935–1972. Springer, 2016.

[18] Gordon Briggs, Tom Williams, Ryan Blake Jackson, and Matthias Scheutz. Why and how robots should say 'no'. *International Journal of Social Robotics*, 14(2):323–339, 2022.

[19] Peter Brown. *The Wild Robot*. Little, Brown Books for Young Readers, 2016.

[20] Adrian Burns, Barry R Greene, Michael J McGrath, Terrance J O'Shea, Benjamin Kuris, Steven M Ayer, Florin Stroiescu, and Victor Cionca. Shimmer™–a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, 10(9):1527–1534, 2010.

[21] Fabio Maria Carlucci, Lorenzo Nardi, Luca Iocchi, and Daniele Nardi. Explicit representation of social norms for social robots. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4191–4196. IEEE, 2015.

[22] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. The robotic social attributes scale (RoSAS) development and validation. In *ACM/IEEE Intl. Conf. on HRI*, pages 254–262, 2017.

[23] Vivienne Bihe Chi and Bertram F Malle. Instruct or evaluate: how people choose to teach norms to social robots. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 718–722. IEEE, 2022.

[24] Adrienne Chung and Rajiv N Rimal. Social norms: A review. *Review of Communication Research*, 4:1–28, 2016.

[25] Robert B Cialdini, Raymond R Reno, and Carl A Kallgren. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *JPSP*, 58(6):1015, 1990.

[26] The Qt Company. Qt for python. https://doc.qt.io/qtforpython/, 2020. Accessed: June 27, 2023.

[27] Stephen Cranefield and Bastin Tony Roy Savarimuthu. Normative multi-agent systems and human-robot interaction. In *Workshop on Robot Behavior Adaptation to Human Social Norms (TSAR)*, 2021.

[28] Kerstin Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences*, 362(1480):679–704, 2007.

[29] Kerstin Dautenhahn, Michael Walters, Sarah Woods, Kheng Lee Koay, Chrystopher L Nehaniv, A Sisbot, Rachid Alami, and Thierry Siméon. How may i serve you? a robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 172–179, 2006.

[30] Kerstin Dautenhahn and Iain Werry. A quantitative technique for analysing robot-human interactions. In *IEEE/RSJ IROS*, volume 2, pages 1132–1138. IEEE, 2002.

[31] Shayn S Davidson, Joseph R Keebler, Tianxin Zhang, Barbara Chaparro, James Szalma, and Christina M Frederick. The development and validation of a universal enjoyment measure: The enjoy scale. *Current Psych.*, pages 1–13, 2022.

[32] Maartje MA De Graaf and Somaya Ben Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and autonomous systems*, 61(12):1476–1486, 2013.

[33] Maartje MA de Graaf, Somaya Ben Allouch, and Jan AGM Van Dijk. Why would I use this in my home? A model of domestic social robot acceptance. *HCI*, 34(2):115–173, 2019.

[34] Maartje MA de Graaf and Bertram F Malle. People's judgments of human and robot behaviors: a robust set of behaviors and some discrepancies. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 97–98, 2018.

[35] Autumn Edwards, Chad Edwards, and Andrew Gambino. The social pragmatics of communication with social robots: Effects of robot message design logic in a regulative context. *International Journal of Social Robotics*, 12(4):945–957, 2020.

[36] Facebook. React - a javascript library for building user interfaces. https://reactjs.org/, 2013. Accessed: June 27, 2023.

[37] Abrar Fallatah, Bohkyung Chun, Sogol Balali, and Heather Knight. "would you please buy me a coffee?" how microcultures impact people's helpful actions toward robots. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 939–950, 2020.

[38] Julia Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International conference on social robotics*, pages 199–208. Springer, 2012.

[39] Kerstin Fischer, Malte Jung, Lars Christian Jensen, and Maria Vanessa aus der Wieschen. Emotion expression in hri–when and why. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 29–38. IEEE, 2019.

[40] Susan T Fiske. *Social beings: Core motives in social psychology*. John Wiley & Sons, 2018.

[41] Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23, 1981.

[42] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003.

[43] Marlena R Fraune, Selma Šabanović, and Takayuki Kanda. Human group presence, group characteristics, and group norms affect human-robot interaction in naturalistic settings. *Frontiers in Robotics and AI*, 6:48, 2019.

[44] Chelsea Frazier-Young, Malcolm McCurry, Kevin Zish, and Greg Trafton. Perceived agency changes performance and moral trust in robots. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44, 2022.

[45] Yotaro Fuse, Hiroshi Takenouchi, and Masataka Tokumaru. A robot in a human–robot group learns group norms and makes decisions through indirect mutual interaction with humans. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 24(1):169–178, 2020.

[46] Yotaro Fuse and Masataka Tokumaru. Navigation model for a robot as a human group member to adapt to changing conditions of personal space. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 24(5):621–629, 2020.

[47] Yotaro Fuse and Masataka Tokumaru. Social influence of group norms developed by human-robot groups. *IEEE Access*, 8:56081–56091, 2020.

[48] Danilo Gallo, Shreepriya Gonzalez-Jimenez, Maria Antonietta Grasso, Cecile Boulard, and Tommaso Colombino. Exploring machine-like behaviors for socially acceptable robot navigation in elevators. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 130–138. IEEE, 2022.

[49] Moojan Ghafurian, Jesse Hoey, and Kerstin Dautenhahn. Social robots for the care of persons with dementia: a systematic review. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(4):1–31, 2021.

[50] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Frontiers in psychology*, 6:931, 2015.

[51] Alain Goudey and Gael Bonnin. Must smart objects look human? study of the impact of anthropomorphism on the acceptance of companion robots. *Recherche et Applications en Marketing (English Edition)*, 31(2):2–20, 2016.

[52] Halfbrick Studios. Fruit Ninja. Mobile application, 2010. Available on: iOS, Android.

[53] Ji Han, Gopika Ajaykumar, Ze Li, and Chien-Ming Huang. Structuring human-robot interactions via interaction conventions. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 341–348. IEEE, 2020.

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

[55] Kerstin Haring, Kristin Nye, Ryan Darby, Elizabeth Phillips, Ewart de Visser, and Chad Tossell. I'm not playing anymore! a study comparing perceptions of robot and human cheating behavior. In *International Conference on Social Robotics*, pages 410–419. Springer, 2019.

[56] Michael Hechter and Karl-Dieter Opp. *Social norms*. Russell Sage Foundation, 2001.

[57] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model. *Intl. J. of Social Robotics*, 2(4):361–375, 2010.

[58] Tessa van der Heiden, Florian Mirus, and Herke van Hoof. Social navigation with human empowerment driven deep reinforcement learning. In *International Conference on Artificial Neural Networks*, pages 395–407. Springer, 2020.

[59] Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, Antti Tammela, Aurélie Clodic, and Rachid Alami. Should a robot guide

like a human? a qualitative four-phase study of a shopping mall robot. In *Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings 11*, pages 548–557. Springer, 2019.

[60] Patrick Holthaus, Catherine Menon, and Farshid Amirabdollahian. How a robot's social credibility affects safety performance. In *International Conference on Social Robotics*, pages 740–749. Springer, 2019.

[61] Shanee S Honig, Tal Oron-Gilad, Hanan Zaichyk, Vardit Sarne-Fleischmann, Samuel Olatunji, and Yael Edan. Toward socially aware person-following robots. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):936–954, 2018.

[62] Christine Horne. *Social norms*, chapter Sociological perspectives on the emergence of norms. In Opp and Hechter [101], 2005.

[63] Ohad Inbar and Joachim Meyer. Politeness counts: perceptions of peacekeeping robots. *IEEE Transactions on Human-Machine Systems*, 49(3):232–240, 2019.

[64] Ákos Jobbágy, Miklós Majnár, Lilla K Tóth, and Péter Nagy. Hrv-based stress level assessment using very short recordings. *Periodica polytechnica Electrical engineering and computer science*, 61(3):238–245, 2017.

[65] Michiel Joosse, Manja Lohse, Niels Van Berkel, Aziez Sardar, and Vanessa Evers. Making appearances: How robots should approach people. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–24, 2021.

[66] Sangmin Kim, Jongsuk Choi, Yoonseob Lim, and Sonya S Kwak. Should a robot follow social norms? human-robot interaction design for social relations in mixed age group. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8234–8239. IEEE, 2022.

[67] Tineke Klamer, Somaya Ben Allouch, and Dirk Heylen. "adventures of harvey"–use, acceptance of and relationship building with a social robot in a domestic environment. In *Human-Robot Personal Relationships: Third International Conference, HRPR 2010, Leiden, The Netherlands, June 23-24, 2010, Revised Selected Papers 3*, pages 74–82. Springer, 2011.

[68] Kheng Lee Koay, Emrah Akin Sisbot, Dag Sverre Syrdal, Mick L Walters, Kerstin Dautenhahn, and Rachid Alami. Exploratory study of a robot approaching a person in the context of handing over an object. In *AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics*, pages 18–24. Stanford, CA, 2007.

[69] Kheng Lee Koay, Michael L Walters, Alex May, Anna Dumitriu, Bruce Christianson, Nathan Burke, and Kerstin Dautenhahn. Exploring robot etiquette: Refining a hri home companion scenario based on feedback from two artists who lived with robots in the uh robot house. In *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings 5*, pages 290–300. Springer, 2013.

[70] Joanna Korman, Anthony Harrison, Malcolm McCurry, and Greg Trafton. Beyond programming: can robots' norm-violating actions elicit mental state attributions? In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 530–531. IEEE, 2019.

[71] Klaus Krippendorff. Bivariate agreement coefficients for reliability of data. *Sociological methodology*, 2:139–150, 1970.

[72] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer, 2008.

[73] Brenda Laurel. Interface agents: Metaphors with character. *Human Values and the design of Computer Technology*, pages 207–219, 1997.

[74] Steven Lawrence, Negin Azizi, Kevin Fan, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. Exploring the nuances: Trust, discomfort, and physiological changes in response to robotic norm violations. Manuscript submitted for publication, 2023.

[75] Steven Lawrence, Negin Azizi, Kevin Fan, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. The impact of social norm violations on participants' perception of and trust in a robot during a competitive game scenario. In *accepted for publication in Proc. IEEE RO-MAN*. IEEE, 2023. In press.

[76] Hee Rin Lee and Selma Sabanović. Culturally variable preferences for robot design and use in south korea, turkey, and the united states. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 17–24, 2014.

[77] Namyeon Lee, Jeonghun Kim, Eunji Kim, and Ohbyung Kwon. The influence of politeness behavior on user compliance with social robots in a healthcare service setting. *International Journal of Social Robotics*, 9:727–743, 2017.

[78] Huao Li, Stephanie Milani, Vigneshram Krishnamoorthy, Michael Lewis, and Katia Sycara. Perceptions of domestic robots' normative behavior across cultures. In

Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pages 345–351, 2019.

[79] Christina Lichtenthäler and Alexandra Kirsch. Legibility of robot behavior: A literature review. 2016.

[80] Christina Lichtenthäler, Annika Peters, Sascha Griffiths, and Alexandra Kirsch. Social navigation-identifying robot navigation patterns in a path crossing scenario. In *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings 5*, pages 84–93. Springer, 2013.

[81] Patrick Lin, Keith Abney, and George A Bekey. *Robot ethics: the ethical and social implications of robotics*. MIT press, 2014.

[82] Torun Lindholm and Sven-Åke Christianson. Intergroup biases and eyewitness testimony. *The Journal of social psychology*, 138(6):710–723, 1998.

[83] Felix Lindner. How to count multiple personal-space intrusions in social robot navigation. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016*, 290:323, 2016.

[84] Alexandru Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati. Evidence that robots trigger a cheating detector in humans. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*, pages 165–172, 2015.

[85] Yao Liu and Shilong Liao. The influence of robot autonomy on perception distance, acceptance and subjective norm. In *Journal of Physics: Conference Series*, volume 2037, page 012116. IOP Publishing, 2021.

[86] Carlos Lobato, Araceli Vega-Magro, Pedro Núñez, and Luis Jesús Manso. Human-robot dialogue and collaboration for social navigation in crowded environments. In *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 1–6. IEEE, 2019.

[87] Joseph B Lyons and Svyatoslav Y Guznov. Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4):440–458, 2019.

[88] Bertram F Malle. *How the mind explains behavior: Folk explanations, meaning, and social interaction*. MIT press, 2006.

[89] Bertram F Malle, Paul Bello, and Matthias Scheutz. Requirements for an artificial agent with norm competence. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 21–27, 2019.

[90] Bertram F. Malle, Eric Rosen, Vivienne B. Chi, Matthew Berg, and Peter Haas. A general methodology for teaching norms to social robots. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1395–1402, 2020.

[91] Bertram F Malle, Eric Rosen, Vivienne B Chi, Matthew Berg, and Peter Haas. A general methodology for teaching norms to social robots. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1395–1402. IEEE, 2020.

[92] Serena Marchesi, Cecilia Roselli, and Agnieszka Wykowska. Cultural values, but not nationality, predict social inclusion of robots. In *International Conference on Social Robotics*, pages 48–57. Springer, 2021.

[93] José M Marques, Vincent Y Yzerbyt, and Jacques-Philippe Leyens. The "black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. *European journal of social psychology*, 18(1):1–16, 1988.

[94] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[95] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and the PRISMA Group*. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, 151(4):264–269, 2009.

[96] Jonathan Mumm and Bilge Mutlu. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 331–338, 2011.

[97] Clifford Nass, Youngme Moon, and Nancy Green. Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of applied social psychology*, 27(10):864–876, 1997.

[98] Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder. Anthropomorphism, agency, and ethopoeia: computers as social actors. In *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*, pages 111–112, 1993.

[99] National Academies of Sciences, Engineering, and Medicine. *Addressing the social and cultural norms that underlie the acceptance of violence: Proceedings of a Workshop—in Brief*. The National Academies Press, Washington, DC, 2018.

[100] Barbara J O'keefe. The logic of message design: Individual differences in reasoning about communication. *Communications Monographs*, 55(1):80–103, 1988.

[101] Karl-Dieter Opp and Michael Hechter, editors. *Social norms*. Russell Sage Foundation, 2005.

[102] Elinor Ostrom. Collective action and the evolution of social norms. *Journal of economic perspectives*, 14(3):137–158, 2000.

[103] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLOS Medicine*, 18(3):e1003583, 2021.

[104] Bjorn Petrak, Katharina Weitz, Ilhan Aslan, and Elisabeth Andre. Let me show you your new home: studying the effect of proxemic-awareness of robots on users' first impressions. In *2019 28th IEEE international conference on robot and human interactive communication (RO-MAN)*, pages 1–7. IEEE, 2019.

[105] Nuno Piçarra, Jean-Christophe Giger, Grzegorz Pochwatko, and Jakub Możaryn. Designing social robots for interaction at work: Socio-cognitive factors underlying intention to work with social robots. *Journal of Automation, Mobile Robotics and Intelligent Systems*, pages 17–26, 2016.

[106] Joelle Pineau, Michael Montemerlo, Martha Pollack, Nicholas Roy, and Sebastian Thrun. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and autonomous systems*, 42(3-4):271–281, 2003.

[107] Andres Pinilla, Jan-Niklas Voigt-Antons, Jaime Garcia, William Raffe, and Sebastian Möller. Real-time affect detection in virtual reality: a technique based on a three-dimensional model of affect and eeg signals. *Frontiers in Virtual Reality*, 3, 2023.

[108] David Porfirio, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. Authoring and verifying human-robot interactions. In *Proceedings of the 31st annual acm symposium on user interface software and technology*, pages 75–86, 2018.

[109] Kun Qian, Xudong Ma, Xianzhong Dai, and Fang Fang. Robotic etiquette: Socially acceptable navigation of service robots with human motion pattern learning and prediction. *Journal of Bionic Engineering*, 7(2):150–160, 2010.

[110] Luis Quintero, John E Muñoz, Jeroen De Mooij, and Michael Gaebler. Excite-o-meter: Software framework to integrate heart activity in virtual reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 357–366. IEEE, 2021.

[111] Alessandra Rossi, Fernando Garcia, Arturo Cruz Maya, Kerstin Dautenhahn, Kheng Lee Koay, Michael L Walters, and Amit K Pandey. Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task. In *Towards Autonomous Robotic Systems: 20th Annual Conference, TAROS 2019, London, UK, July 3–5, 2019, Proceedings, Part I 20*, pages 349–361. Springer, 2019.

[112] Amalia G Sabiescu and Nemanja Memarovic. Participatory design for cultural representation: A cultural transparency perspective. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV 14*, pages 611–618. Springer, 2013.

[113] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pages 141–148, 2015.

[114] Nicole Salomons, Sarah Strohkorb Sebo, Meiying Qin, and Brian Scassellati. A minority of one against a majority of robots: Robots cause normative and informational conformity. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(2):1–22, 2021.

[115] Eduardo Benítez Sandoval, Jürgen Brandstatter, Utku Yalcin, and Christoph Bartneck. Robot likeability and reciprocity in human robot interaction: Using ultimatum game to determinate reciprocal likeable robot strategies. *International Journal of Social Robotics*, 13:851–862, 2021.

[116] Morito Sato, Masahiko Mikawa, Makoto Fujisawa, and Wasuke Hiiragi. Social norm based collision avoidance in human-robot coexistence environment. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 3801–3806. IEEE, 2018.

[117] Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

[118] Shimmer Sensing. ConsensysPro Software. https://shimmersensing.com/product/consensyspro-software/, 2023. Accessed on: June 6, 2023.

[119] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. No fair!! an interaction with a cheating robot. In *ACM/IEEE Intl. Conf. on HRI*, pages 219–226. IEEE, 2010.

[120] Felix Wilhelm Siebert, Jacobe Klein, Matthias Rötting, and Eileen Roesler. The influence of distance and lateral offset of follow me robots on user perception. *Frontiers in Robotics and AI*, 7:74, 2020.

[121] Jesper Simonsen and Toni Robertson. *Routledge international handbook of participatory design*, volume 711. Routledge New York, 2013.

[122] Kiran Jot Singh, Divneet Singh Kapoor, and Balwinder Singh Sohi. Understanding socially aware robot navigation. *Journal of Engineering Research*, 10(1A):131–149, 2022.

[123] Andrew Steain, Christopher John Stanton, and Catherine J Stevens. The black sheep effect: The case of the deviant ingroup robot. *PloS one*, 14(10):e0222975, 2019.

[124] Dag Sverre Syrdal, Kerstin Dautenhahn, Michael L Walters, and Kheng Lee Koay. Sharing spaces with robots in a home scenario-anthropomorphic attributions and their effect on proxemic expectations and evaluations in a live hri trial. In *AAAI fall symposium: AI in Eldercare: new solutions to old problems*, volume 2008, pages 116–123, 2008.

[125] Debora Viana Thompson, Rebecca W Hamilton, and Roland T Rust. Feature fatigue: When product capabilities become too much of a good thing. *Journal of marketing research*, 42(4):431–442, 2005.

[126] Elena Torta, Raymond H Cuijpers, James F Juola, and David van der Pol. Design of robust robotic proxemic behaviour. In *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3*, pages 21–30. Springer, 2011.

[127] Gabriele Trovato, Tatsuhiro Kishi, Nobutsuna Endo, Kenji Hashimoto, and Atsuo Takanishi. A cross-cultural study on generation of culture dependent facial expressions of humanoid social robot. In *ICSR*, pages 35–44, 2012.

[128] Pauline Trung, Manuel Giuliani, Michael Miksch, Gerald Stollnberger, Susanne Stadler, Nicole Mirnig, and Manfred Tscheligi. Head and shoulders: Automatic error

detection in human-robot interaction. In *Proceedings of the 19th ACM international conference on multimodal interaction*, pages 181–188, 2017.

[129] Katherine M Tsui, Munjal Desai, and Holly A Yanco. Considering the bystander's perspective for indirect human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 129–130. IEEE, 2010.

[130] Daniel Ullman and Bertram F. Malle. The multidimensional measure of trust (MDMT), v1. Technical report, Brown University, Providence, RI, Available at http://research.clps.brown.edu/SocCogSci/Measures/MDMT_v1.pdf, 2018.

[131] Dieter Vanderelst, Corinne Jorgenson, Ali I Ozkes, and Jurgen Willems. Are robots to be created in our own image? testing the ethical equivalence of robots and humans. *International Journal of Social Robotics*, pages 1–15, 2022.

[132] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.

[133] Joey AF Verhoeven and Peter AM Ruijten. Gesture cues in navigational robots. In *International Conference on Social Robotics*, pages 378–387. Springer, 2019.

[134] Veritas Health Innovation. Covidence systematic review software. Melbourne, Australia. Available at https://www.covidence.org, 2023.

[135] Francesco Vigni and Silvia Rossi. Exploring non-verbal strategies for initiating an hri. In *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I*, pages 280–289. Springer, 2023.

[136] Michael L Walters, Kerstin Dautenhahn, Sarah N Woods, and Kheng Lee Koay. Robotic etiquette: results from user studies involving a fetch and carry task. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 317–324, 2007.

[137] Lin Wang, Pei-Luen Patrick Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. When in rome: The role of culture & context in adherence to robot recommendations. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 359–366. IEEE, 2010.

[138] Alicja Wasik, Stevan Tomic, Alessandro Saffiotti, Federico Pecora, Alcherio Martinoli, and Pedro U Lima. Towards norm realization in institutions mediating human-robot societies. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 297–304. IEEE, 2018.

[139] Ruchen Wen, Zhao Han, and Tom Williams. Teacher, teammate, subordinate, friend: Generating norm violation responses grounded in role-based relational norms. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 353–362. IEEE, 2022.

[140] Yalun Wen, Xingwei Wu, Katsu Yamane, and Soshi Iba. Socially-aware mobile robot trajectories for face-to-face interactions. In *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I*, pages 3–13. Springer, 2023.

[141] Melonee Wise, Michael Ferguson, Derek King, Eric Diehr, and David Dymesich. Fetch and freight: Standard platforms for service robot applications. In *Workshop on autonomous mobile service robots*, 2016.

[142] Ben Wright, Magdalena Bugajska, William Adams, Ed Lawson, J Malcolm McCurry, and J Gregory Trafton. Proxemic reasoning for group approach. In *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings 12*, pages 48–60. Springer, 2020.

[143] Shannon Yasuda, Devon Doheny, Nicole Salomons, Sarah Strohkorb Sebo, and Brian Scassellati. Perceived agency of a social norm violating robot. In *Proceedings of the annual meeting of the cognitive science society*, 2020.

[144] Mateusz Żarkowski. Multi-party turn-taking in repeated human–robot interactions: an interdisciplinary evaluation. *International Journal of Social Robotics*, 11(5):693–707, 2019.

[145] Min Zhao, Dan Li, Zhun Wu, Shiyan Li, Xiaonan Zhang, Lu Ye, Guangfu Zhou, and Daisong Guan. Stepped warm-up–the progressive interaction approach for human-robot interaction in public. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*, pages 309–327. Springer, 2019.

[146] Biwen Zhu and David Kaber. Effects of etiquette strategy on human–robot interaction in a simulated medicine delivery task. *Intelligent Service Robotics*, 5:199–210, 2012.

[147] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics*, 7:347–360, 2015.

APPENDICES

Appendix A

Ethics Approval

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

---

Principal Investigator: Kerstin Dautenhahn (Electrical and Computer Engineering)

Student investigator: Steven Lawrence (Math Faculty Computing Facility)

Collaborator: Melanie Jouaiti (Electrical and Computer Engineering)

Faculty supervisor: Chrystopher Nehaniv (Electrical and Computer Engineering)

Student investigator: Kevin Fan (Electrical and Computer Engineering)

Student investigator: Austin Kothig (Systems Design Engineering)

File #: 44436

Title: Robot scavenger hunt

The Human Research Ethics Board is pleased to inform you this study has been reviewed and given ethics clearance.

Initial Approval Date: 06/28/22 (m/d/y)

University of Waterloo Research Ethics Boards are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. Both Boards are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREB) and IRB00007409 (CREB).

This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 06/29/23 (m/d/y)

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Board