

Quantifying Human Mental States in Physical Human-Robot Interaction

by

Nourhan Abdulazeem

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2024

© Nourhan Abdulazeem 2024

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

To support independent living among older adults, we propose the introduction of robots as domestic service providers capable of assisting with daily activities like showering, toileting, and transferring. These tasks often require physical interaction, necessitating robots with efficient physical capabilities and a nuanced social understanding of humans, such as stress levels and attitudes toward the robot (robot perception). Our objective is to explore potential real-time (objective) quantification approaches in physical Human-Robot Interaction (pHRI), with the ultimate aim of developing a human mental state predictor to enhance robots' decision-making abilities.

We identify cognitive ergonomics and belief as two key human factors associated with cognitive aspects that can sufficiently represent the human mental state. To accurately measure these dimensions, we recognize the need for a combination of objective and subjective approaches. However, literature examining the suitability of subjective measures in pHRI, which is crucial for obtaining reliable objective quantification, is limited. Therefore, our research includes an assessment of a critical subjective quantification tool for a key human factor, i.e., robot perception, followed by an exploration of correlations between subjective and objective measures.

Our findings reveal significant correlations between subjective and objective measures, indicating the potential for objective indicators such as skin temperature and task performance to reliably infer cognitive ergonomics and belief in collaborative pHRI scenarios. Furthermore, we observe promising associations between pre-gathered data, such as personality traits and demographics, and measurable dimensions of cognitive ergonomics and belief, suggesting a supplementary approach for enhancing inference using objective measures.

Acknowledgements

I express my deepest gratitude to my supervisor, Prof. Yue Hu. I consider myself profoundly fortunate to have earned Prof. Hu's trust and unwavering faith in both me and my vision, which have empowered me to pursue this thesis with confidence. Her invaluable support nurtured my growth and allowed me to flourish according to my aspirations. I am thankful for her dedication to providing me with diverse research experiences, and especially grateful for her patience in guiding me through the challenges and milestones of this endeavor. Prof. Hu's ability to embrace my mistakes and instill in me the belief that setbacks and misfortunes are integral parts of the learning process has been truly invaluable.

I extend my sincere appreciation to my colleagues and friends at the A.I.R (Active and Interactive Robotics) Lab for their abiding support and assistance throughout the execution of my experiments, as well as during all mock presentations, where they provided me with their honest and invaluable feedback.

Last but not least, I am indebted to my parents, sisters, and lifelong friends for their unwavering support and encouragement. Their belief in me and willingness to stand by me through moments of stress, anxiety, and frustration have empowered me to pursue my goals with unwavering determination, continually progressing forward.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgements	iv
List of Figures	viii
List of Tables	xi
List of Abbreviations	xii
1 Introduction	1
1.1 Publications and Roadmap	2
1.2 Background	3
2 Literature Review	5
2.1 Introduction	5
2.2 Existing Review Papers	6
2.2.1 Literature Gap	7
2.3 Research Questions	9
2.4 Methodology	11
2.4.1 Search Strategy	11
2.4.2 Study Selection and Quality Assessment	12
2.5 Results	14
2.5.1 Mapping Scheme for pHRI	14
2.5.2 Human Factors Classification	20
2.5.3 Quantification Approaches of Human Factors	29
2.6 Discussion	31
2.6.1 Mapping Scheme for pHRI	31

2.6.2	Human Factor Classification	36
2.6.3	Human Factors Quantification	39
2.7	Limitations of the review	41
2.8	Conclusion	43
2.9	Aligning with Thesis Objective	44
3	The Impact of Robot’s Facial Emotional Expressions on Light Physical Exercise	45
3.1	Introduction	45
3.2	Related works	47
3.2.1	Facial expressions in HRI	47
3.2.2	Age groups differences	48
3.3	Research Questions	49
3.4	Methodology	51
3.4.1	Experiment design	51
3.4.2	Participants	53
3.4.3	Procedure	55
3.4.4	Measures	56
3.5	Results	57
3.5.1	Variance tests	57
3.5.2	Correlation Analysis	61
3.6	Discussion	66
3.6.1	Participant perceptions	66
3.6.2	Comparative analysis of older and younger adult age groups	69
3.7	Limitations	70
3.8	Conclusion	72
4	Assessment of Robot Perception Subjective Quantification Approach	73
4.1	Introduction	73
4.2	Method	74
4.3	Results	74
4.4	Discussion	77
4.5	Aligning with Thesis Objective	78
5	Quantifying Human Mental State in Interactive pHRI: Maintaining Balancing	79
5.1	Introduction	79
5.2	Related Work	81
5.3	Research Questions	83

5.4	Methodology	84
5.4.1	Equipment	85
5.4.2	Robot Control	85
5.4.3	Measures	87
5.4.4	System Design	90
5.4.5	Participants	93
5.4.6	Procedure	94
5.4.7	Data Processing	96
5.5	Results	97
5.5.1	Variance tests	101
5.5.2	Correlations	101
5.5.3	Answers to RQs	101
5.6	Discussion	103
5.7	Conclusion	105
6	Conclusion	107
6.1	Future Work	110
	References	112
	APPENDICES	132
A	Questionnaires	133
A.1	StuMMB-Q	133

List of Figures

1.1	Example of an indirect physical Human-Robot Interaction (pHRI) application: a collaborative robotic arm mediates the interaction between the user and an object, such as handing over a book.	4
1.2	Example of a direct pHRI application: a user exercises with a collaborative robotic arm by pushing against its stiff joints.	4
2.1	Most representative measurable dimensions in the pHRI literature according to the results of the literature review. This diagram is adapted to pHRI from the holistic conceptual model adapted for Human-Robot Interaction (HRI) proposed in [36]. Measurable dimensions are the human factors that can be quantified and interpreted, and theoretical constructs are abstract concepts that encompass various measurable dimensions.	21
2.2	Number of studies emerged in each human factor type	23
2.3	Number of studies emerged in each measurable dimension	23
2.4	Number of studies that evaluated each human factor type in each direct pHRI category.	24
2.5	Number of studies that evaluated each human factor type in each indirect pHRI category	26
2.6	Number of studies relied on each quantification approach	29
2.7	Quantification approaches developed in literature. MM stands for Mathematical Model, ML stands for Machine Learning, Q stands for Questionnaire, and PS stands for Physiological Signals.	31
2.8	Number of studies relied on each assessment approach in each human factor type	34
3.1	Participant performing a physical pushing exercise with Sawyer displaying a facial expression on the flat screen panel.	46
3.2	Sawyer SDK logo displayed as Inactive Facial Expression	52
3.3	Sawyer facial emotional expressions [53], 3.3b is used as the facial expression for the Unresponsive Facial Expression condition.	52

3.4	Boxplots illustrate older adults 3.4a, younger adults 3.4b, and combined all participants' 3.4c perceived competence, discomfort, warmth, safety, and objective engagement. The ratio E_{obj} is represented as a value potentially exceeding 1. Competence, perceived safety, discomfort, warmth, and safety ratings are normalized to fall within the range of 0 to 1. Significance levels ($** := p - value < 0.01$) are indicated on lines between conditions.	58
3.5	Bar plots illustrating older adults, younger adults, and combined all participants' perceived trust and engagement for each condition on a 5-point Likert Scale (1 = highest, 5 = lowest).	59
3.6	Heatmap illustrates Spearman's correlations observed for each condition for the older adults age group. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05, < 0.01$, and < 0.001 with a $ \rho > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.	63
3.7	Heatmap illustrates Spearman's correlations observed for each condition for the younger adults age group. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05, < 0.01$, and < 0.001 with a $ \rho > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.	64
3.8	Heatmap illustrates Spearman's correlations observed for each condition for the all the participants . Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05, < 0.01$, and < 0.001 with a $ \rho > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.	65
4.1	Scree plots of the factor analysis performed on perceived safety (Godspeed) for the inactive condition only based on the Kaiser-Meyer-Olkin (KMO) values, and the three dimensions of Robot Social Attribute Scale (RoSAS) (competence, warmth, and discomfort), for all three conditions.	76
5.1	Interactive pHRI experiment setting.	86
5.2	Boxplots illustrate participants' personality traits according to Ten Item Personality Inventory (TIPI). Each personality trait is normalized to fall within the range of 0 to 1.	98

5.3	Boxplots depict participants' objective stress and robot perception measures, including skin temperature growth in °C/s and the percentage difference in skin temperature, alongside cognitive load measures represented by the performance index and overall performance indicated by the total imbalance durations in seconds across the actual session. Sign of percentage differences in skin temperature indicates whether there is an increase or decrease compared to baseline.	98
5.4	Boxplots illustrate participants' subjective cognitive load, encompassing mental load, and mental effort. In addition to usability ratings. Each measurable dimension ratings are normalized to fall within the range of 0 to 1. Significance levels ($* := p - value < 0.05$) is indicated on lines between conditions.	99
5.5	Boxplots depict participants' subjective perceptions of the robot, including competence, discomfort, warmth, and perceived safety. Each measurable dimension ratings are normalized to fall within the range of 0 to 1.	99
5.6	Heatmap illustrates Spearman's correlations observed for the passive condition between demographics, personality traits, physical data, physiological data, and measurable dimensions representing participants' mental states. Correlations marked with '*', and '**' indicate a $p - value < 0.05$, and < 0.01 , respectively.	100
5.7	Heatmap illustrates Spearman's correlations observed for the active condition between demographics, personality traits, physical data, physiological data, and measurable dimensions representing participants' mental states. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05$, < 0.01 , and < 0.001 , respectively.	102

List of Tables

2.1	Mapping scheme for pHRI studies based on experimental scenarios addressing RQ1 and presenting the characteristics of each set of studies under each subcategory.	16
2.1	<i>Cont.</i>	17
2.2	Common studies identified within the subcategories of the two proposed categories of direct pHRI: purpose and duration.	18
2.3	A brief description of some measurable dimensions.	22
2.4	Human factor types, their respective measurable dimensions, and the studies that have emerged in each measurable dimension, contributing to the answer of RQ2	25
2.4	Human factor types and measurable dimensions addressed according to the proposed mapping scheme for pHRI, with a focus on studies proposing quantification approaches.	27
2.4	<i>Cont.</i>	28
2.5	Quantification approaches employed within each measurable dimension, highlighting studies proposing quantification approaches. Studies using mathematical and machine learning models are categorized based on the input requirements of each model, including task parameters, physical data, and physiological signals.	32
2.5	<i>Cont.</i>	33
3.1	Physical Exercise pHRI experiment sample description for Older Adults . .	54
3.2	Physical Exercise pHRI experiment sample description for Younger Adults	54
5.1	Interactive pHRI experiment sample description	94

List of Abbreviations

BA Bridge Amplifier [85](#)

GSR Galvanic Skin Response [83](#), [85](#), [96](#)

HRI Human-Robot Interaction [viii](#), [5](#), [9](#), [10](#), [18](#), [20](#), [21](#), [34](#), [36](#), [38–40](#), [42](#), [43](#), [72–74](#), [78](#), [104](#), [105](#), [108–110](#)

ICSR International Conference of Social Robotics [3](#)

IMU Inertial Measurements Unit [85](#), [86](#), [90](#), [91](#), [94–96](#)

KMO Kaiser-Meyer-Olkin [ix](#), [74–76](#), [78](#)

NASA-TLX NASA Task Load Index [44](#), [89](#)

pHRI physical Human-Robot Interaction [viii](#), [ix](#), [xi](#), [2–5](#), [7–11](#), [13](#), [18–21](#), [24](#), [26–31](#), [34–40](#), [43–50](#), [54](#), [66–70](#), [72–74](#), [78–84](#), [86](#), [87](#), [89](#), [94](#), [104–111](#)

PPG Photoplethysmography [44](#), [80](#), [85](#), [86](#), [88](#), [94–96](#)

ROS Robot Operating System [92](#)

RoSAS Robot Social Attribute Scale [ix](#), [44](#), [56](#), [73](#), [74](#), [76–78](#), [89](#), [105](#), [108](#)

sHRI Social Human-Robot Interaction [45](#), [47](#)

SNS Sympathetic Nervous System [87](#)

StuMMB-Q Students' Mental Load and Mental Effort in Biology Education-Questionnaire [89](#), [97](#), [104](#), [108](#), [133](#)

SUS System Usability Scale [89](#), [100](#)

TPI Ten Item Personality Inventory [ix](#), [89](#), [94](#), [98](#), [101](#)

Chapter 1

Introduction

As the global population continues to age rapidly due to increased life expectancy and declining birth rates, the demand for healthcare services and caregivers for older adults is growing [129]. In Japan, over 30% of the population is already over 60, while it is projected that by 2050, one in four individuals in North America and Europe could be aged 65 or older [128]. From 2015 to 2050, the proportion of the global population aged over 60 years is estimated to almost double from 12% to 22% [129].

As a result, there is an increasing demand for solutions that enhance the quality of life and independence of older adults to address the looming shortage of caregivers [3]. To achieve this, the Society 5.0 vision aims to introduce robots as domestic service providers to assist with daily living activities and create novel values in healthcare and caregiving domains [81]. Given that caregiving is closely associated with physical interaction tasks in daily activities such as showering, eating, toileting, and transferring [154], robots introduced into these application domains must be capable of efficient physical support. It is noteworthy that physical support does not involve physical assistance only, but also emotional support through physical comfort such as patting or hugging.

To ensure the successful integration of robots into these fields, a practical approach would be to equip them with the capability to provide multiple services instead of limiting them to a single task (e.g. a robot that only cleans the floor, or only performs feeding), i.e., *robotic multiple service providers*. The success of this approach has been demonstrated by the widespread of cellphones and computers, which seamlessly integrate numerous functionalities, such as cameras, internet connectivity, multimedia playback, health and fitness tracking, and mobile payments. The popularity of multi-functional devices underscores a preference among consumers for maximizing productivity and minimizing the need for

multiple specialized devices. Additionally, it promotes synergy and interoperability within its operational environment [19].

A crucial characteristic of multiple service robots is the possession of dexterity and *manipulation capabilities*, as it enables them to perform various tasks efficiently, similar to how humans operate. By equipping robots with such versatility, they are endowed the capability for handling diverse domestic scenarios, ranging from showering and feeding to the co-manipulation of heavy equipment and assisting with sit-and-stand tasks. This adaptability positions them as valuable assets in enhancing daily living experiences and supporting individuals across a spectrum of activities.

Furthermore, a seamless assimilation of robots would require them to develop a social understanding of humans, which will help robots make decisions according to their human partners' state, such as their level of anxiety, emotional state, physical comfort, social distancing, or mental workload. This is anticipated to endow robots with the ability to engage with humans in a socially acceptable manner. As a first step towards achieving this goal, robots need to possess the capability of predicting the human state, i.e. *quantification of human factors*. It is worth mentioning that this goal is part of a comprehensive vision to develop adaptive control frameworks that can make optimal decisions by taking into account not only the human state but also environmental conditions and a collaborative task goal.

To this end, this thesis aims to explore potential real-time quantification approaches, i.e., objective quantification approaches, for human mental states in [pHRI](#). The ultimate goal is to contribute to the development of a human mental state predictor, which will enhance robots' decision-making capabilities in various tasks.

1.1 Publications and Roadmap

To establish the research foundation, we delved into existing literature on approaches to quantify human mental states. This exploration aimed to address fundamental questions: 'What is human mental state?' and 'How can we quantify human mental state?' Chapter 2 emerged from this endeavor. Our findings were subsequently published as a journal paper in *Sensors* in August 2023 [1]. This literature review identified crucial gaps necessary to address if we intend to integrate robots as domestic service providers for daily living activities. Notably, two significant gaps were identified: (1) the lack of validation for commonly used subjective quantification approaches in [pHRI](#) contexts; and (2) the under-explored area of objective quantification approaches.

We commenced our research by evaluating a commonly used subjective approach to quantify robot perception—a key human factor in [pHRI](#) [1]. Evaluating subjective approaches in different experimental scenarios is essential to establish their reliability. Only when a subjective approach is deemed reliable can its correlation with objective quantification approaches yield trustworthy results, which are essential for informing advanced control decisions. We conducted a user study in a light physical exercise context, as presented in Chapter 3, to acquire the necessary data for assessing subjective tools reliability across different age groups (younger adults and older adults) with a relatively large sample size. Chapter 4 details the process and results of evaluating the reliability of these subjective tools. The results for the older age group were published in the [International Conference of Social Robotics \(ICSR\) 2023](#) [2]. The findings from the entire sample, encompassing both age groups, along with the assessment approach, will be submitted as a journal paper to the *International Journal of Social Robotics*, the primary venue for this type of research.

Having confirmed robot perception subjective approach reliability, we progressed to explore objective quantification approaches in [pHRI](#). As detailed in Chapter 5, our focus was on studying various physical robot behaviors within a balancing context through a user study. This approach enabled us to uncover meaningful correlations between objective and subjective quantification methods. The potential of this research extends to at least two additional publications, and the acquired data holds significant value for further analysis. It is crucial to emphasize that each user study conducted stands as an independent investigation, tailored to fulfill the objectives of this thesis.

Chapter 6 summarizes the objectives of this thesis and offers suggestions regarding critical gaps in the literature that require further investigation.

1.2 Background

The definitions of [pHRI](#) and related human factors in the literature are often ambiguous. The literature on human factors in human-robot interaction (HRI) is characterized by great diversity in its usage and a lack of agreement on unified definitions or terminologies, as noted in [166]. For this work, we define *human factors* as:

Data drawn from humans that can quantifiably interpret the human state to be used in a human-in-the-loop control framework.

Examples of human factors are trust, anxiety, mental workload, physical comfort, and perceived safety. The human state refers to the degree or level of a particular human

factor. For instance, if we consider trust as a human factor, its absence or low levels would be referred to as the human state of 'lack of trust.' As far as the authors are aware, no previous reviews have focused on investigating the studies that have specifically addressed measurable human factors interpreting human state in the context of **pHRI**.

Similar to the ambiguity surrounding the definition of human factors, the field of **pHRI** also grapples with the difficulty of reaching a consensus on a unified definition. In [146], the authors define **pHRI** as the domain that focuses on the physical aspects of the interaction between humans and robots in a shared workspace, including factors such as robot speed and distance. Whereas in [134], **pHRI** referred to a broad area of applications that include cooperation, assistance, teleoperation, and entertainment. In [155], the authors categorized the physical connection between humans and robots as either proximally, where there is an exchange of forces between parties, i.e., physical coupling, or remotely, referring to teleoperation. Further restrictions on the area that **pHRI** refers to are adopted by [101], where authors considered only physical coupling (direct physical contact) between humans and robots, including those mediated by an object. Building upon the perspective of the field presented in [101] on **pHRI**, we define **indirect pHRI** as physical coupling that is mediated by an object, shown in Figure 1.1 and **direct pHRI** as physical coupling that occurs directly between parties without object mediation, shown in Figure 1.2.



Figure 1.1: Example of an indirect **pHRI** application: a collaborative robotic arm mediates the interaction between the user and an object, such as handing over a book.



Figure 1.2: Example of a direct **pHRI** application: a user exercises with a collaborative robotic arm by pushing against its stiff joints.

Chapter 2

Literature Review

2.1 Introduction

First and foremost, to aid in the development of multiple service robots, it is necessary to identify research efforts that have: 1) examined quantifiable and measurable factors that can represent the human state, 2) addressed efficient physical interaction, and 3) exploited robots with strong manipulation capabilities. To this end, our goal in this chapter, serving as a literature review, is to identify studies that have quantifiably assessed **human factors** in [pHRI](#) scenarios, with a specific focus on robots that are capable of *manipulation*, meaning those equipped with multi-degree-of-freedom arm(s).

To gain a comprehensive understanding of our topic of interest, it is imperative to discern research gaps as well as evaluate the advancements achieved in the literature. In pursuit of this, we adopted a novel holistic conceptual model for human factors classification in [HRI](#) [36], which enabled us to provide an unbiased overview of the most commonly studied human factors and the approaches used to quantify them. However, to facilitate deeper analysis and enhance our understanding of these factors, we first identified the various contexts in which these human factors have been investigated in [pHRI](#), which can serve as a potential mapping scheme for the field. To establish this mapping scheme, we have incorporated Petersen et al. [138] suggestions for a topic-specific classification of studies. These classifications may arise from the mapping study itself or be drawn from existing literature. While this review does not aim to provide a strict systematic literature review, it offers an objective summary of the progress made in human factors in [pHRI](#) over the last 15 years. We specifically selected studies involving physical coupling between humans and robots, encompassing both direct and indirect [pHRI](#). The robots used in

these studies possess manipulation capabilities, such as multi-degrees of freedom arm(s). Moreover, these studies focus on evaluating human factors during physical interaction. However, we excluded studies that involved physical coupling with robots for purposes other than collaboration (e.g., rehabilitation), as well as those assessing human factors in individuals with mental disorders or in physical interaction with baby-like robots. These exclusions were made as they do not align with the primary goal of this review. The purpose of this summary is to assist in accelerating the development of a streamlined adaptive control framework for multiple service robots.

2.2 Existing Review Papers

Remarkable efforts have been expended to comprehensively review HRI studies pertaining to human factors. Hopko et al. [74] conducted a systematic literature review, where they assessed the most frequently studied human factors and quantification methods used in industrial shared spaces. Their findings revealed that trust, cognitive workload, and anxiety were the most commonly investigated human factors. While Hopko et al. dealt with perceived safety as a distinct human factor, Rubagotti et al. [146] incorporated related human factors, such as trust, anxiety, and comfort, as sub-factors under the umbrella of perceived safety. Their primary objective was to understand how physical factors, such as distance and robot speed, and different types of robots, such as industrial manipulators, mobile robots, and humanoids, can influence the perception of safety. Similarly, Ortenzi et al. [130] took a holistic approach to address different human factors by utilizing "user experience" as an umbrella term instead of examining each factor separately. Their study focused on reviewing efforts made in handovers between humans and robots. Likewise, Prasad et al. [139] examined previous studies in handshake scenarios to identify the most commonly addressed human factors. Coronado et al. [36] conducted a literature review to provide an overview of the quantification methods used in HRI industrial applications to evaluate performance and measurable dimensions of human factors. Their work resulted in a taxonomy of performance aspects and a holistic conceptual model for the most commonly addressed human factors in HRI. Their review also revealed that, after physical safety, trust was the most frequently addressed human factor that influences human performance. Simone et al. [165] studied the impact of human-robot collaboration on human factors in industrial settings. Their review highlighted that successful human-robot interaction depends on several crucial factors, including trust, usability, and acceptance. Additionally, they concluded that the interaction can affect stress and mental workload either positively or negatively. Lorenzini et al. [100] conducted a narrative review to identify the most effective assessment methods for physical and cognitive ergonomics in workplace environ-

ments. Their review focused on available monitoring technologies that endow robots with the ability to adapt online to the mental states and physical needs of workers.

Notably, each of the aforementioned reviews found that questionnaires were the predominant assessment approach utilized in studies across diverse scenarios.

2.2.1 Literature Gap

Evidently, a dearth of research has specifically examined human factors in [pHRI](#) scenarios or investigated the studies that quantified the impact of human factors within domestic contexts. Moreover, past attempts to classify human factors in this domain have faced numerous challenges, necessitating the use of various methodological workarounds. In light of these gaps, we endeavor to provide an objective overview of the progress made in this field, in order to promote the development of adaptive control frameworks for multiple service robots. Efficiently predicting the human state is a crucial first step in this development.

To this end, our review distinguishes itself by proposing a comprehensive mapping scheme for [pHRI](#) applications and experimental scenarios, considering human factors in [pHRI](#), adopting a human factors model with the aim of enhancing quantification approaches and highlighting potential relationships between measurable dimensions, and organizing the quantification data-based approaches based on their inputs.

Numerous reviews have implemented methodological workarounds to address the challenge of diverse terminologies used to describe various human factors. Some authors, such as Rubagotti et al. [146] have employed the term "perceived safety", while others, like Ortenzi et al. [130], have used the broader concept of "user experience" to encompass multiple human factors, without delving into specific classifications. Other reviews, like Hopko et al. [74], addressed individual human factors but did not provide clear definitions or outlines.

Similarly, classifying human factors has been a major challenge in this review due to the lack of systematic or standardized definitions in HRI literature, making it difficult to determine the most commonly addressed human factor. For example, some studies use terms like stress [76], strain [55], and frustration [63] without providing clear definitions, making it inaccurate to group them together under one umbrella term. Likewise, legibility [38], and predictability [70] of robot's motion; and physical ergonomics [188], physical comfort [38], fatigue [136], health and safety [115] are two sets of examples of different human factors that overlap but cannot be lumped under a single comprehensive term within the set. Additionally, perceived enjoyment [63], and enjoyability [50]; and psychological safety, mental safety, and subjective safety [146] are also examples of human factors that

are distinct but interconnected. We anticipate that this may have caused an unintentional research bias in favor of human factors that have limited synonyms like trust. Trust has been identified as a common human factor in several HRI reviews and has even been given special attention in dedicated reviews [94].

Furthermore, the challenge of assessing progress in human factors in HRI is exacerbated when studies used generic terms like preference [34] and enjoyment [50] without providing clear definitions for each term. This highlights the importance of defining and classifying human factors in a more precise and standardized way to avoid unintentional research bias. To address this challenge, we explored various classification schemes, such as dividing human factors into cognitive and physical ergonomics, as proposed by Lorenzini [100], or further categorizing them into load and perceptual factors. However, we found that human factors in pHRI, which involved both domestic and industrial scenarios, are more complex and diverse than these schemes can account for, as they may involve emotional factors in addition to cognitive and physical ones.

Therefore, we chose to build upon the holistic conceptual model proposed by Coronado et al. [36] for human factors in HRI. By adopting this model, we aim to develop a standardized approach that can systematically assess the progress of human factors research in HRI, including pHRI scenarios.

Human Factors Conceptual Model

Coronado et al. [36] employed a systematic approach to identifying quality factors, measures, and metrics from the HRI literature. The researchers developed two models: one for classifying performance-related aspects and another for focusing on human-centered aspects of robotics systems. Although initially designed for industrial and collaborative robotics, the authors encouraged the adaptation of these models to other robotics disciplines. Consequently, their human-centered model, known as the 'human factors holistic conceptual model,' was adopted for our field of interest, pHRI. The proposed holistic conceptual model adapted for HRI, based on [151], reduces confusion between the concepts of usability and user experience in HRI literature. This model illustrates the relationships among usability, user experience, hedonomics, and ergonomics, and identifies four distinct human factor types: cognitive ergonomics, physical ergonomics, belief, and affect. Below we will briefly describe the definition of each concept.

User experience, according to the ISO 9241-11:2018 (ergonomics of human-system interaction) [80], includes "all user's emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors, and accomplishments that occur before, during, and after use". Whereas, *usability* is considered as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency,

and satisfaction in a specified context of use”.

Ergonomics, which is classified into physical and cognitive ergonomics in this review, aims to mitigate injuries, pain, and suffering, while *hedonomics*, which is classified into belief and affect in this review, focuses on the pleasant and enjoyable aspects of interactions.

Physical ergonomics encompasses the potential adverse effects on the human body during an interaction, such as postures, repetitive movements, heavy workloads, or forces. *Cognitive ergonomics* is concerned with designing systems that align with the perceptual and psychological abilities of users. *Affect* pertains to emotional-related terms, whereas *belief* refers to cognitive responses that can trigger emotions (i.e., affective responses).

By utilizing this conceptual model, we anticipate addressing the deficiency of unified definitions of human factors by classifying them under one of the four human factor types. Following the naming convention presented in Coronado et al., we use the term *measurable dimensions* to refer to human factors that can be quantified and measured, such as trust, anxiety, and fatigue. In contrast, we use *human factor types* to refer to broader categories of human factors, such as cognitive ergonomics, physical ergonomics, belief, and affect.

It is important to clarify that while the terms *human factors* and *ergonomics* are sometimes used interchangeably in the [pHRI](#) literature (especially in the context of human-robot collaboration), our adopted conceptual model differentiates between the two. In our review, we use *human factors* and *measurable dimensions* interchangeably (see the definition of measurable dimensions in the previous paragraph). On the other hand, *ergonomics* serves as the broader umbrella term, encompassing various human factors and measurable dimensions, with a specific focus on the safety and functionality aspects of the interaction. This distinction is evident from Hancock’s Hedonomic pyramid [67], which is the basis of the adopted conceptual model.

We are confident that this approach will facilitate the investigation of relationships between measurable dimensions within each human factor type. This will allow us to gain a more nuanced understanding of how different aspects of human factors interact with each other.

2.3 Research Questions

Although the literature on [HRI](#) has addressed human factors considerations, there has been a notable lack of dedicated attention to [pHRI](#) as a specific area of study requiring a structured analysis of trends and developments. In this review paper, we aim to identify

the different applications or experimental scenarios in which human factors have been studied in the field, with a specific emphasis on those that involve robots with manipulation capabilities. Therefore, by creating a mapping scheme, we can help researchers navigate and better understand the field. As a result, we formulated the following research question:

RQ1: What are the existing applications or experimental scenarios in [pHRI](#) that consider human factors, and how can they be categorized?

To address this question, we will examine the applications and experimental scenarios that researchers have utilized in their experiments for both direct and indirect [pHRI](#).

In order to identify the most common human factor type addressed in [pHRI](#) studies that involve robots with manipulation capabilities, we formulated the following research question:

RQ2: Which of the four human factor types (belief, affect, physical ergonomics, and cognitive ergonomics) is most frequently investigated to develop quantitative measures of the human state?

We hypothesize that belief will be the answer to this question, as it encompasses trust as one of its measurable dimensions, which is consistently identified as one of the most frequently addressed human factors in [HRI](#).

Quantification of human factors based on measurable data plays a key role in the development of adaptive control frameworks that can effectively predict human states and make decisions based on them. As a result, we have formulated the following research question:

RQ3: What are the quantification approaches (questionnaire, physical data, mathematical models, physiological signals, and machine learning models), employed to evaluate and analyze the impact of human factors in [pHRI](#)?

To identify the quantification approaches based on measurable data, we will categorize the gathered studies into two categories based on the objectives of the studies that addressed human factors. We call this categorization *Human Factor Usage* which includes impact and quantification. The *impact* category explores the effect of manipulating robot parameters on human factors, i.e. the impact of those manipulations, while the *quantification* category will explore proposed quantification approaches for human factors or

their correlation with measurable entities such as physical data (e.g., body posture and interaction forces) and physiological data. We hypothesize that, based on the literature, questionnaires will be the most commonly used approach for quantifying human factors in [pHRI](#). Our objective is to identify promising quantification techniques that have been introduced or adapted specifically for [pHRI](#).

2.4 Methodology

The methodology used to conduct this review follows the guidelines proposed by Petersen et al. [138], which were adapted to the context of HRI by Coronado et al. [36]. The methodology consists of several steps:

1. Justifying the need for conducting a literature review, as described in Section 2.1, Introduction.
2. Defining the research questions, as explained in Section 2.3, Research Questions.
3. Determining a search strategy, assessing its comprehensiveness, establishing selection criteria, and conducting a quality assessment. These steps are detailed in this section.
4. Extracting data from the identified studies, organizing and categorizing the information obtained, and visualizing the results. These steps are detailed in Section 2.5, Results.
5. Addressing the limitations of the review by conducting a validity assessment, as described in Section 2.7, Limitations of the Review.

2.4.1 Search Strategy

Given that [pHRI](#) is a relatively young field and the overall population is not well-known beforehand, conducting a comprehensive database search for human factors is challenging. Unlike more established fields, the literature on [pHRI](#) is not yet fully developed, which makes it harder to determine suitable search terms. Additionally, as noted earlier, there is a lack of standardization in the reporting of human factors in HRI, which further complicates the search process. To address this challenge, we employed a combination of search strategies to generate a suitable set of keywords for a comprehensive and effective search string.

As a first step, a combination of manual search [87] and snowballing [180] was used. We identified relevant studies from the references and citations of the following HRI reviews that specifically addressed human factors: [74], [146], [163], [164], and [165]. In addition to examining the proceedings of the following 2022 conferences which were not indexed yet at the time of writing this manuscript: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE International Conference on Intelligent Robots and Systems (IROS), and IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO). Furthermore, we included some prominent papers based on expert recommendations.

Initially, keywords were extracted from the set of papers obtained from the previous step, which formed the initial set of papers. However, it was found that barely half of them used the word "physical" in any of their search fields, despite the physical coupling occurring in the interaction. Moreover, the most common words detected in the papers were too generic for a database search, such as "human-robot interaction". Additionally, the term "human factors" was not commonly used in the papers, as researchers would not necessarily use it when evaluating factors such as trust [66, 94], anxiety [75], or other related concepts [119].

The search string was developed based on combining the generic term detected and the human factors terms identified in the initially considered studies, such as comfort, cognitive load, mental load, fatigue, stress, anxiety, fear, acceptance, perception, emotions, trust, safety, ergonomics, and hedonomics. The string was refined by excluding topics covered under the broad domain of HRI, such as wearables, teleoperation, telepresence, rehabilitation, and autonomous vehicles. The search was conducted on Scopus and Web of Science search engines, and the resulting papers were evaluated using a test-set approach. Each iteration was assessed by checking for the presence or absence of the initial set of papers obtained from the manual and snowballing search approaches. We ended up with 711 papers, after the removal of duplicates.

2.4.2 Study Selection and Quality Assessment

The study selection process involved the following steps:

1. Initial screening: Abstracts of papers were checked against the inclusion criteria, with full texts read for papers that were in doubt. This resulted in 128 papers.
2. Full-text assessment: Quality assessment was conducted during this phase to ensure

that each paper included information relevant to the research questions. Exclusion criteria were applied, resulting in 103 papers being included in the study.

The following points summarize the characteristics of each study included in this review, i.e., inclusion criteria:

1. Physical coupling between humans and robots takes place, direct or indirect [pHRI](#).
2. Robot used has manipulation capabilities (i.e. has multi degrees of freedom arm(s)).
3. Human factors are evaluated during physical interaction.

While our primary objective is to focus on [pHRI](#) in domestic scenarios, we did not discard studies in industrial settings, as the field of [pHRI](#) is still young, and studying empirical data from various contexts can provide valuable insights.

Although some papers met the inclusion criteria, they did not directly contribute to the main objectives of this study. As a result, the following exclusion criteria were established:

1. Physical coupling with robots for intents other than collaboration [[54](#), [118](#)], including rehabilitation.
2. Human factor evaluations of individuals with mental disorders were excluded [[150](#)], as experimental verification is required to generalize the factors of these special populations to the neurotypical population.
3. Unclear identification of physical coupling between agents within a study.
4. Manifestly missing methodology of human factors assessment or analysis [[47](#), [104](#)].
5. Physical interaction with baby-like robots, since this type of robot is not expected to contribute to the main objective of this study [[90](#), [182](#)].
6. Studies that are not in English [[79](#)].
7. Inaccessible full text.

2.5 Results

The following subsections present the results of our analysis, which aim to address the research questions we formulated. We first present a mapping scheme for pHRI in Subsection 2.5.1, followed by the classification of human factors based on our conceptual model in Subsection 2.5.2. Then, in Subsection 2.5.3, we identify the most frequently addressed human factors, and examine both the commonly employed quantification approaches and the proposed quantification approaches.

2.5.1 Mapping Scheme for pHRI

Due to the vast applications of direct and indirect pHRI categories, it is crucial to provide subcategories that facilitate a more detailed understanding and easier navigation of the field. To address this need, we developed a mapping scheme for pHRI that categorizes the gathered studies based on the common experimental scenarios identified in direct and indirect pHRI, as shown in Table 2.1, where each subcategory is described. These subcategories are based on the specific characteristics of the studies in each category, thereby addressing **RQ1**.

Direct pHRI

Direct pHRI has received considerably more attention in human factor studies than indirect pHRI, with over two-thirds of the studies conducted in this area focusing on direct pHRI (70 studies) and only 35 studies on indirect pHRI.

The subcategories for direct pHRI proposed in this study are based on the interaction context, including the *purpose* and *duration* of contact. These classifications provide a preliminary overview of the research in this area. As the field matures, these subcategories can be further refined to provide a more detailed understanding.

Table 2.1 shows that inconsistent contact has been explored in more studies than consistent contact, while the topic of non-functional contact has been studied almost twice as often as the topic of functional contact.

Many of the studies examined in the context of inconsistent contact were also included in the context of non-functional contact, such as studies that have investigated the NAO robot’s reactive behavior in response to social touch, including one conducted by [126]. Another study, by [110], compared the instrumental and affective touch of the NAO robot to that of a Stretch robot in a caregiving scenario. Moreover, several studies have focused

on the use of NAO in caregiving scenarios that involve inconsistent and non-functional touch, such as those by [60, 4] and [8]. These findings are consistent with several reviews in the HRI field that have highlighted the potential of NAO in caregiving scenarios. Fitter and Kuchenbecker have conducted research on different clapping contexts using a Baxter robot. Their studies include teaching [51], and playing with the robot [49] [50]. Meanwhile, Shiomi et al. used a female-looking Android to investigate the effects of its subtle reactions to touch [158][157] and the impact of pre-touch reaction distance on humans [160].

Additionally, Hu et al. investigated the impact of unexpected physical actions applied by a Sawyer robot in a gaming context with the aim of helping participants accomplish higher scores [76, 75].

In contrast, many studies in the consistent contact category were also included in the functional touch category. Some of those studies are conducted in dancing context [132]; direct manipulation of the arm to follow certain trajectory [61] [28] [175] [41] [181], different from kinesthetic teaching; leading robots with mobile platforms using their manipulators in a nursing context [32] and a social context [46]. Fewer studies were identified as common between the consistent contact and non-functional touch categories. These studies primarily focused on hugging and handshake scenarios with humanoids. Some studies on handshake scenarios were conducted with the Meka robot [171] and HRP-2 robot [14], while other studies investigated kinesthetic teaching using a 7-DoF Franka Emika manipulator [10]. Studies conducted on hugging scenarios investigated the impact of manipulating visual and tactile appearance using an ARMAR-IIIb robot [170], the duration and style of hugs using a CASTOR robot [20], and the impact of the perceived gender using a Metahug system [156] [157].

The least common studies were found between the inconsistent contact and functional touch categories. Many of these studies were conducted in a nursing scenario, where the impact of instrumental and affective touch was investigated using the Cody robot [33] and NAO robot [110]. Additionally, the NAO robot was used to investigate the impact of touch style in a nursing scenario [22], and gaze height and speech timing before and after touches were examined. However, few studies have been conducted in an industrial setting, with the majority focusing on the impact of tapping a robotic manipulator as an indicator to execute a certain command [109] [148].

Table 2.1: Mapping scheme for **pHRI** studies based on experimental scenarios addressing **RQ1** and presenting the characteristics of each set of studies under each subcategory.

Categories		Characteristics	Studies	Number of Studies
Direct pHRI	Purpose	Non-Functional Touch	Includes studies that involve physical touch between the robot and the human with the intention of communicating a psychological state, such as social touch [126], or for the purpose of exploration or curiosity, as in [84].	45
		Functional Touch	Includes studies that involved physical contact between the robot and the human for a specific purpose, such as manipulation, assistance, or control, i.e., instrumental touch. Figure 1.2 shows an example of functional touch in a physical exercise scenario.	26
	Duration	Consistent Contact	Includes studies requiring continuous direct contact throughout the interaction.	29
		Inconsistent Contact	Includes studies where direct contact is not necessary throughout the entire interaction.	40

Table 2.1: *Cont.*

	Categories	Characteristics	Studies	Number of Stud- ies
Indirect pHRI	Assembly	Includes studies where one agent holds a part while the other agent assembles it with another part.	[168 , 112 , 44 , 6 , 183 , 27 , 26]	7
	Handover	Includes studies where one agent is handing over an object to the other agent. Figure 1.1 shows an example of a handover scenario.	[66 , 5 , 133 , 93 , 38 , 92 , 65 , 167 , 188 , 140 , 70 , 124 , 35 , 77 , 183]	15
	Co-manipulation	Includes studies where both human and robot agents manipulate an object in the environment with the goal of changing its position or orientation.	[121 , 117 , 177 , 112 , 39 , 99 , 86 , 183]	8
	Atypical	Includes studies that do not fall into any of the other categories of indirect pHRI , such as assistive holding/drilling or dressing assistance.	[136 , 135 , 137 , 88 , 48 , 57 , 31 , 183 , 105]	9

It is worth noting that among the studies that examined the impact of touch on human factors, only two studies ([[110](#)] and [[33](#)]) explored the effects of functional and non-functional touch in [HRI](#), both of which were conducted in a nursing context and categorized under inconsistent contact. No studies have examined the effects of consistent versus inconsistent contact on human factors in direct [pHRI](#).

Table [2.2](#) shows the common studies between all subcategories of direct [pHRI](#), including functional touch, non-functional touch, consistent contact, and inconsistent contact.

This table has been compiled due to the noticeable presence of shared studies among the subcategories. It is important to note that the classification of direct [pHRI](#) is currently considered a proposal, primarily because the field is relatively young compared to indirect [pHRI](#). This is evident from the predominance of exploratory experiments conducted in direct [pHRI](#), such as gaming and clapping studies, with only a few applications having emerged, such as handshaking and hugging. Unlike indirect [pHRI](#), which allows for clearer classification based on common applications, direct [pHRI](#)'s diverse and evolving nature makes such categorization challenging at this stage.

Table 2.2: Common studies identified within the subcategories of the two proposed categories of direct [pHRI](#): purpose and duration.

	Functional Touch	Non-Functional Touch
Consistent Contact	[175 , 76 , 41 , 111 , 181 , 75 , 132 , 152 , 143 , 96 , 61 , 39 , 28 , 123 , 46 , 32 , 183 , 141]	[14 , 186 , 156 , 56 , 20 , 157 , 179 , 170 , 171 , 7]
Inconsistent Contact	[109 , 148 , 110 , 108 , 159 , 22 , 33 , 10]	[110 , 9 , 21 , 126 , 69 , 84 , 60 , 95 , 187 , 185 , 184 , 158 , 50 , 49 , 35 , 178 , 160 , 127 , 113 , 73 , 51 , 11 , 4 , 8 , 107 , 83 , 68 , 33 , 120 , 82 , 43 , 37 , 122]

Indirect [pHRI](#)

In contrast to direct [pHRI](#), indirect [pHRI](#) has clear classifications based on real-world applications like assembly, handover, co-manipulation, and atypical scenarios. These subcategories can be useful for researchers seeking to understand the potential use cases and challenges of [HRI](#) in different practice settings. As shown in Table [2.1](#), there has been a particular focus on addressing human factors in handovers, which has received the most attention compared to other forms of indirect [pHRI](#) applications.

Handover scenarios in industrial settings have been the subject of several studies. For instance, some investigations have used Universal Robot (UR) manipulators to compare comfortable handovers to those that encourage dynamic behaviors [[188](#)], while others have employed UR to explore scenarios where a manipulator is completing a preliminary task and an operator requests the handover of a tool through natural communication, such as voice commands [[93](#)]. Similarly, some studies have examined industrial setting that incorporate

handovers using a Panda manipulator [92]. Other research has focused on determining the most ergonomic tool orientation for handovers [140] [5]. Various handover parameters have also been explored, including physical signs of apprehension [65], the initial position of the robot arm before handover, the robot’s grasp method, and the retraction speed of the arm following handover [133].

In the context of domestic scenarios, there has been limited research on the impact of robot parameters on human factors during handovers. One study examined comfortable robot-to-human handovers in a caregiving scenario for older adults, using a Pepper robot [124]. Another investigation explored affective and efficient interaction styles in a cooking context, where both faulty and successful handovers were performed during the cooking task [66].

In a general context where handovers may occur in domestic and industrial environments, a study investigated handovers using a mobile manipulator that transferred objects to humans without stopping [70].

In the atypical category, numerous studies have focused on industrial applications of assistive holding, drilling, and cutting, using manipulators [105] [31] [135] [137] [48]. A few studies have explored the potential of robots as dressing assistants [57]. Another study investigated the repositioning of patients using robots in a caregiving scenario [88].

Co-manipulation occurs in both domestic and industrial scenarios and is usually investigated in similar ways. It typically involves investigating ways to reduce joint overload when humans collaborate with robots in lifting heavy objects [39, 112, 117, 99, 86]. Some studies also focus on mutual adaptation in this collaboration [121], as well as human intention and behavior [177].

Studies in the assembly category are typically conducted in an industrial setting due to the nature of the application. Some of these studies have explored the use of two robotic arms to hold an object for an operator while they work on it [6]. Other studies have focused on typical assembly tasks where a robot holds an object and the user assembles another part into it [44, 112] [26, 27], or welds a wire into it [168].

Only one study in the indirect **pHRI** category, conducted in an industrial setting, investigated both assembly and co-manipulation tasks [112]. This finding suggests that more research is needed in this area to gain a better understanding of the challenges and opportunities of the indirect **pHRI** field, since many industrial applications may require more than one type of indirect **pHRI**.

Among the studies reviewed, some studies investigated both direct and indirect forms of **pHRI**. For instance, one study incorporated social touch and handover, respectively, into a collaborative task of building a towering toy [35]. The combination of different types

of physical interaction research shows promising directions for addressing contexts where social behavior and physical assistance are combined.

2.5.2 Human Factors Classification

To address **RQ2**, we adopted the holistic conceptual model of human factors in [HRI](#) proposed by Coronado et al. [36], as explained in Section 2.2, and adapted it to demonstrate the relationships between the most relevant measurable dimensions found in the [pHRI](#) literature. Figure 2.1 illustrates the relationships between usability, user experience (UX), hedonomics, and ergonomics, based on Coronado et al. definitions, which are discussed in Section 2.3.

Although the conceptual model used for human factor types does not fully resolve the issue of inconsistent terminologies or definitions in the field, grouping measurable dimensions, such as trust, robot perception, and anxiety, into subcategories (human factor types), such as belief, can help address this issue. This is because presenting the trends of a category instead of individual human factors can reduce bias that arises from the usage of synonyms or variations in the ways researchers refer to human factors.

On a lower level, to address the lack of unified terminologies in the literature, this study adopts a strategy of lumping some human factors under one measurable dimension, as described in Table 2.3. To avoid ambiguity for the reader, a brief description of some dimensions is provided. The adopted [HRI](#) conceptual model allows for some overlap between dimensions within a human factor type, as it is difficult to define sharp outlines for each dimension. The number of times a dimension is used, whether it is lumped under a broader one, detected individually, or intersects with another in definition, will not affect the overall result of the addressed frequency of a human factor type. However, if two or more human factor types share a dimension, their overall results will be biased. Therefore, in this review, all dimensions considered for each human factor type do not intersect over more than one type. If a dimension intersects over multiple types, it is reviewed and divided into more specific dimensions to avoid bias.

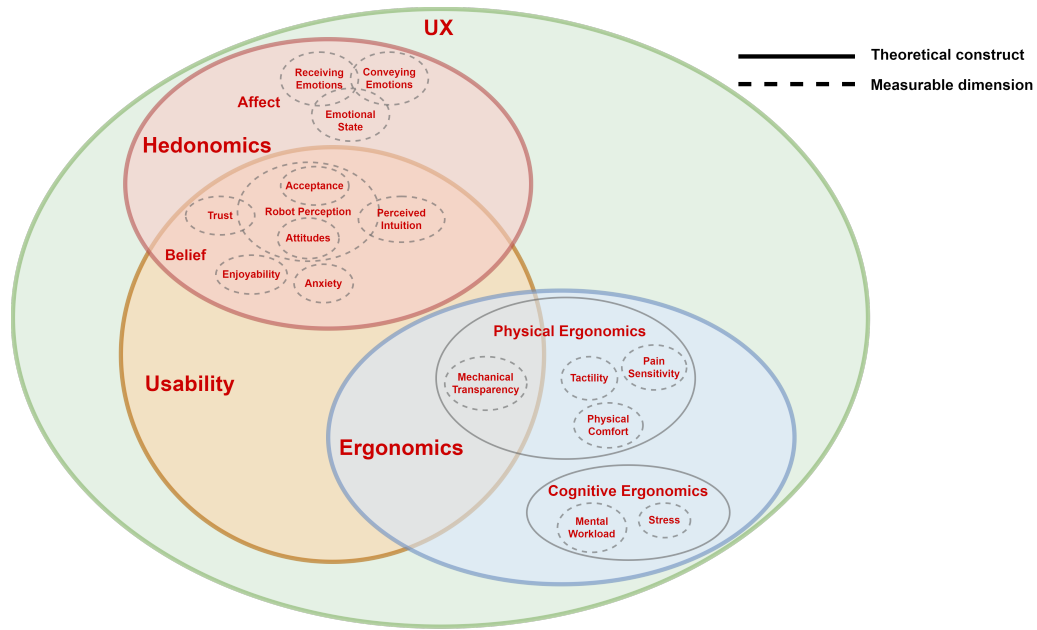


Figure 2.1: Most representative measurable dimensions in the pHRI literature according to the results of the literature review. This diagram is adapted to pHRI from the holistic conceptual model adapted for HRI proposed in [36]. Measurable dimensions are the human factors that can be quantified and interpreted, and theoretical constructs are abstract concepts that encompass various measurable dimensions.

Table 2.3: A brief description of some measurable dimensions.

Measurable Dimension	Description
Tactility	Indicates the perceived pleasantness when touching a robot.
Physical Comfort	Includes studies that have evaluated human posture, muscular effort, joint torque overloading, peri-personal space, comfortable handover, legibility, and physical safety.
Mechanical Transparency	“Quantifies the ability of a robot to follow the movements imposed by the operator without noting any resistant effort” [117]. It includes the predictability of the robot’s motion in following user physical instructions, naturalness and smoothness of the motion, sense of being in control, responsiveness to physical instruction of participants, feeling of resistive force, and frustration.
Robot Perception	Indicates the user’s perception toward the robot. It includes attitudes, impressions, opinions, preferences, favorability, likeability, willingness for another interaction, behavior perception, politeness, anthropomorphism, animacy, vitality, perceived naturalness, agency, perceived intelligence, competence, perceived safety, emotional security, harmlessness, toughness, familiarity, friendship, companionship, friendliness, warmth, psychological comfort, helpfulness, reliable alliance, acceptance, ease of use, and perceived performance.
Perceived Intuition	Includes goal perception, whether the robot understands the goal of the task or not, robot intelligence, willingness to follow the robot’s suggestion, dependability, understanding of robot intention, and perceived robot helpfulness.
Conveying Emotions	Indicates humans’ perspective on how they should convey their emotions to robots by physical touch.
Receiving Emotions	Indicates humans’ perspective of how humans expect to receive a robot’s emotions through physical touch.
Emotional State	Indicates recognition of a human’s emotional state during interaction without necessarily conveying their emotions using physical touch.

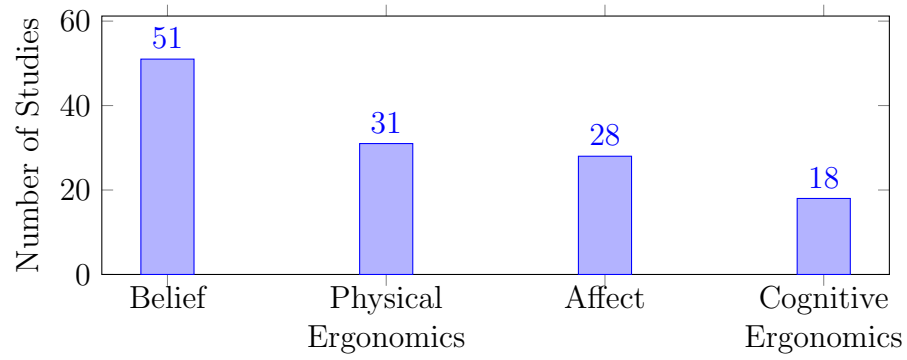


Figure 2.2: Number of studies emerged in each human factor type

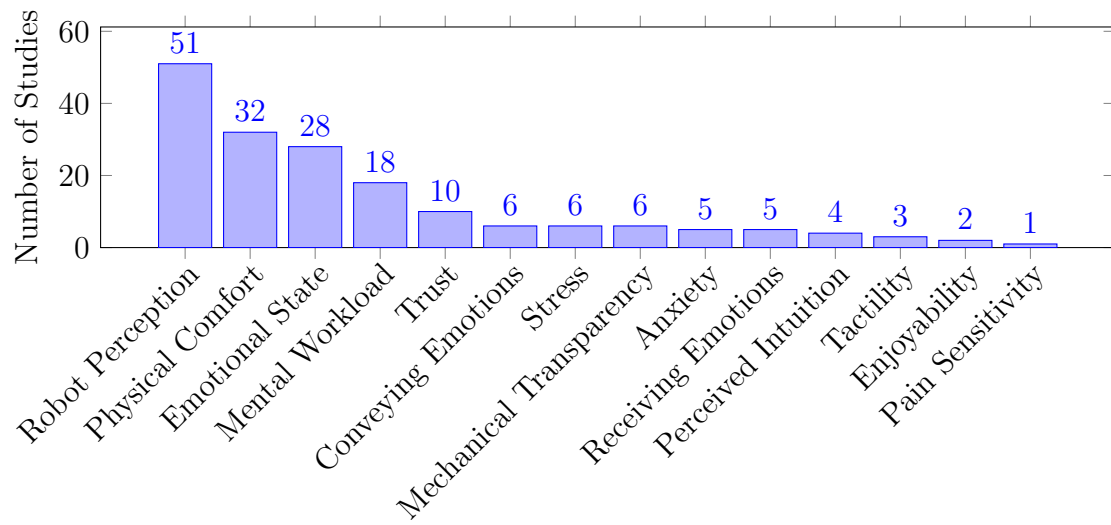


Figure 2.3: Number of studies emerged in each measurable dimension

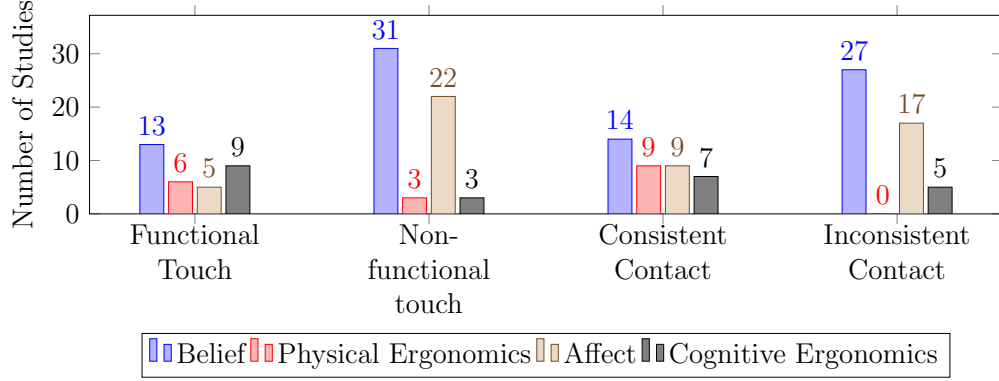


Figure 2.4: Number of studies that evaluated each human factor type in each direct pHRI category.

In order to identify the most commonly addressed human factors, i.e., address **RQ2**, we present two sets of figures: Figures 2.2 and 2.3 show the emerged human factor types and measurable dimensions across all gathered studies, with Table 2.4 showing the studies detected in each human factor type. It is notable how belief and the measurable dimension, robot perception, have drawn the most attention in the pHRI field. This is reflected in Figure 2.4 to a certain extent, which shows the number of studies that examined human factor types in different direct pHRI categories. Analogous to the considerable overlap between the non-functional touch and inconsistent contact categories, Figure 2.4 shows that both categories addressed hedonomics more than ergonomics. Similarly, the functional touch and consistent contact categories exhibited a similar trend, with both categories emphasizing belief more than other human factor types. However, Figure 2.5, which illustrates the number of studies that examined human factor types within subcategories of indirect pHRI, shows that physical ergonomics is the leading human factor type in these scenarios. Similarly, physical ergonomics is the only aspect that receives attention in the atypical category. Likewise, cognitive ergonomics and affect were not addressed in the co-manipulation category. Table 2.4 shows the studies detected in each measurable dimension and pHRI type.

During the analysis of the results, we observed that several studies evaluated two human factor types simultaneously within the same experiment. Notably, the highest number of studies focused on the evaluation of belief and affect, including [126] [6] [69] [175] [178] [51] [33] [41] [75] [91] [105]. Following this, several studies explored the relationship between cognitive ergonomics and belief, such as [6] [50] [51] [75] [76] [175] [66] [92] [156]. Fewer number of studies examined cognitive ergonomics with physical ergonomics, including [6] [175] [188] [38] and physical ergonomics with belief, such as [25] [6] [175] [157]. Lastly, it was

Table 2.4: Human factor types, their respective measurable dimensions, and the studies that have emerged in each measurable dimension, contributing to the answer of **RQ2**.

Human Factor Type	Measurable Dimension	Studies
Cognitive Ergonomics	Mental Workload	[6, 50, 51, 32, 123, 75, 76, 108, 111, 175, 66, 92, 93, 188, 10]
	Stress	[156, 22, 75, 76, 38, 92]
Physical Ergonomics	Pain Sensitivity	[17]
	Tactility	[25, 179, 157]
	Physical Comfort	[48, 57, 39, 31, 99, 86, 175, 5, 38, 167, 188, 140, 70, 124, 112, 136, 135, 88, 137, 183, 26, 27, 141]
	Mechanical Transparency	[6, 39, 28, 46, 117, 105]
Belief	Robot Perception	[14, 126, 69, 60, 25, 156, 95, 158, 159, 50, 178, 160, 113, 73, 11, 170, 83, 68, 33, 120, 82, 43, 37, 77, 122, 41, 75, 76, 21, 132, 66, 133, 110, 91, 148, 109, 152, 61, 56, 20, 157, 105]
	Trust	[60, 95, 6, 120, 43, 175, 65, 92, 121, 177]
	Perceived Intuition	[6, 51, 37, 75]
	Enjoyability	[50, 51]
	Anxiety	[122, 41, 75, 76, 56]
Affect	Emotional State	[69, 44, 184, 49, 35, 178, 127, 51, 96, 33, 41, 75, 181, 91, 168, 20, 21, 170]
	Conveying Emotions	[126, 69, 8, 4, 107, 9]
	Receiving Emotions	[186, 187, 171, 7, 185]

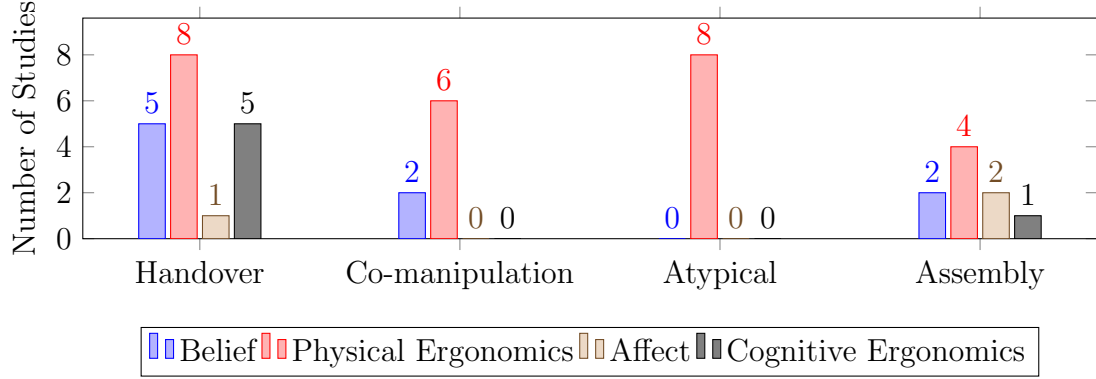


Figure 2.5: Number of studies that evaluated each human factor type in each indirect pHRI category

found that the simultaneous consideration of affect and cognitive ergonomics within the same experiment was relatively rare. Only a few studies, such as [51] and [75], specifically explored this combination. Notably, there were no studies found that examined measurable dimensions from physical ergonomics and affect in the same experiment.

Additionally, some studies, such as Amanhoud et al. [6] and Wang et al. [175], assessed three human factor types at once, namely cognitive ergonomics, physical ergonomics, and belief. Similarly, Fitter et al. [51] and Hu et al. [75] investigated the effects of cognitive ergonomics, belief, and affect.

Table 2.4: Human factor types and measurable dimensions addressed according to the proposed mapping scheme for **pHRI**, with a focus on studies proposing quantification approaches.

Human Factor Type	Measurable Dimension	Direct pHRI				Indirect pHRI			
		Purpose		Duration		Assembly	Handover	Co-Manipulation	Atypical
		Functional Touch	Non-Functional Touch	Consistent	Inconsistent				
Cognitive Ergonomics	Mental Work-load	[175, 111, 123, 75] *	[50, 51]	[75, 111, 123, 10, 175] * [32, 76]	[108, 50, 51]	[6]	[92] * [66, 93, 188]		
	Stress	[75] * [76]	[156]	[156]	[22]	[92] * [38]			
Physical Ergonomics	Pain Sensitivity			[17] *					
	Tactility		[25, 179, 157]	[179, 157]					
	Physical Comfort	[175, 183] * [141]		[175, 183] * [141]		[183, 26, 27] * [112]	[124, 183] * [5, 38, 167, 140, 70, 188]	[39, 99, 86, 183] * [112]	[135, 136, 48, 137, 57, 31, 183] * [88]
	Mechanical Transparency	[39, 28, 46]		[39, 28, 46]		[6]		[117]	[117, 105]

* Studies that have proposed quantification approaches.

Table 2.4: *Cont.*

Human Factor Type	Measurable Dimension	Direct pHRI				Indirect pHRI			
		Purpose		Duration		Assembly	Handover	Co-Manipulation	Atypical
		Functional Touch	Non-Functional Touch	Consistent	Inconsistent				
Belief	Robot Perception	[75, 152] *	[14, 126, 68, 122] * [69, 60, 25, 156, 95, 158, 50, 178, 160, 113, 73, 11, 170, 83, 33, 120, 82, 43, 37, 21, 110, 91, 84, 56, 20, 157]	[75, 14] * [175, 76, 181, 132, 186, 156, 143, 56, 20, 96, 179, 41]	[126, 69, 68, 122] * [148, 110, 91, 21, 95, 158, 159, 50, 178, 160, 113, 73, 11, 83, 33, 120, 82, 43, 37, 84, 60, 51]		[133] * [66, 77]	[105]	
	Trust	[175] *	[60, 95, 120, 43]		[60, 95, 120, 43]	[6]	[65, 92] *	[177] * [121]	
	Perceived Intuition	[75] *	[51, 37]	[75] *	[51, 37]	[6]			
	Enjoyability		[50, 51]		[50, 51]				
	Anxiety	[75] * [76]	[122] * [56]	[75] * [41, 76]	[122] * [56]				
Affect	Emotional State	[75] * [41, 33, 181, 96]	[69, 127] * [91, 184, 49, 35, 178, 51, 33, 21, 170]	[75] * [41, 170, 76]	[69, 127] * [91, 184, 49, 35, 178, 51, 33, 21]	[44] * [168]	[35]		
	Conveying Emotions		[126, 69, 8, 107] * [9, 4]		[126, 69, 8, 107] * [4, 9]				
	Receiving Emotions		[186] * [187, 171, 7, 185]	[171, 7]	[187, 185]				

* Studies that have proposed quantification approaches.

2.5.3 Quantification Approaches of Human Factors

Less than a third of the studies included in this analysis correlated dimensions with measurable entities or proposed quantification approaches for different human factors, as revealed by the human factor usage classification. The analysis identified 83 papers in the impact category and 32 studies in the quantification category. Some studies validated their proposed quantification approaches by manipulating robot parameters and observing their impact on the human state, including studies by [31], [39], [65], [69], [75], [92], [111], [123], [124], [126], [133], [135], [137], [183], [26], [26]. As a result, these studies were also included in the impact category, which contained more studies than the quantification category. Table 2.4 includes the studies that emerged in each human factor usage category, organized according to human factor types and measurable dimensions.

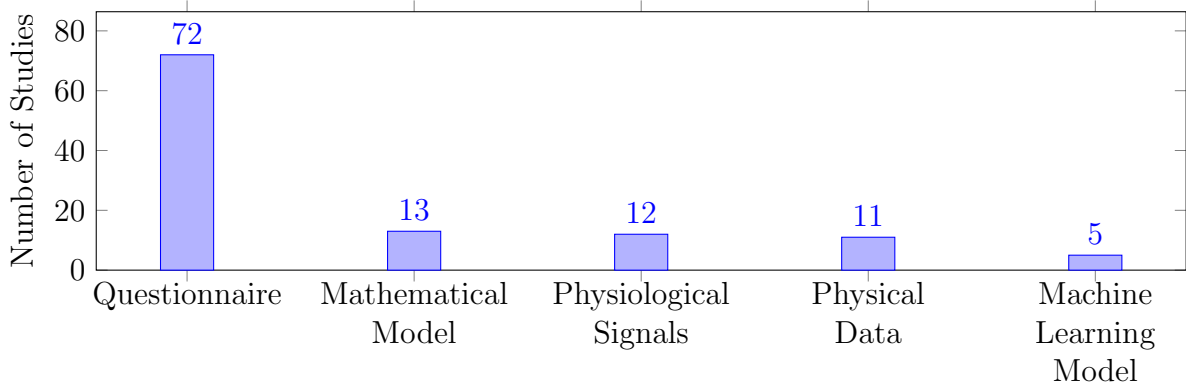


Figure 2.6: Number of studies relied on each quantification approach

Confirming the hypothesis for **RQ3**, questionnaires were found to be the most commonly used approach for human factors quantification in **pHRI**, as shown in Figure 2.6. However, the remaining approaches were used with much lower frequencies compared to questionnaires. To determine the role of questionnaires in **pHRI**, either proposed as a quantification approach or employed as an assessment approach, further analysis was conducted. Despite the fact that questionnaires were heavily employed for assessment, Figure 2.7 indicates that only three studies proposed and/or validated standardized questionnaires for **pHRI**. QUEAD, a questionnaire for evaluating physical assistive devices, was developed and evaluated in a direct **pHRI** scenario where participants solve a maze by guiding a laser pointer attached to a robotic arm’s end-effector [152]. Other efforts focused on validating standardized questionnaires in different scenarios. For instance, the Robot Social Attribute Scale (RoSAS) questionnaire [29] is validated for indirect **pHRI** in a handover scenario [133].

On the other hand, the proposed quantification approaches heavily relied on physical data (e.g., force, pressure, body posture, gaze direction, etc.), either through direct correlation or mathematical and machine learning models. Table 2.5 shows the studies that employed quantification approaches for each measurable dimension and highlighted the studies that proposed quantification approaches. For a comprehensive understanding and easier navigation of the quantification approaches in the field, we organized the studies that utilized mathematical models and machine learning models in Tables 2.4 and 2.5 based on the input requirements of each model, including task parameters, physical data, and physiological signals.

Consistent with the overall trend, the questionnaire approach was found to be the leading method for evaluating human factors in all human factor types, as depicted in Figure 2.8. Although belief was heavily assessed using questionnaires and physiological signals, no studies developed machine learning models to quantify any of its measurable dimensions. Cognitive ergonomics did not see any studies that addressed physical data, while in affect there was a moderate reliance on physiological signals.

It is worth mentioning that the initial steps towards developing models capable of predicting human states during pHRI involved studies that adopted both mathematical and machine learning models. For instance, in [123], researchers utilized physiological data and task parameters to estimate human workload and effort during a gaming scenario. Three modalities were measured and analyzed: electroencephalography (EEG), skin conductance, respiration, and eye tracking. Their findings revealed that task parameters such as task score and movement velocity provided a more accurate estimate of workload and effort than the physiological modalities.

Another notable example is the online mathematical model proposed in [92], which incorporates gaze direction, head pose, self-touching occurrences, and body language to evaluate mental effort and stress levels in an industrial setting involving assembly and handover interactions. The authors utilized questionnaires and physiological data to test their experiment hypothesis and validated their mathematical model with the same data.

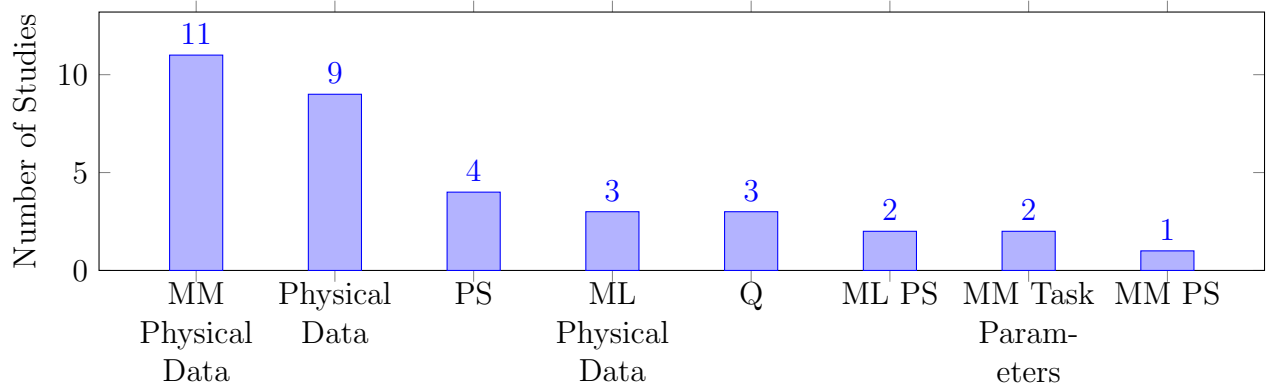


Figure 2.7: Quantification approaches developed in literature. MM stands for Mathematical Model, ML stands for Machine Learning, Q stands for Questionnaire, and PS stands for Physiological Signals.

2.6 Discussion

In light of the literature gaps identified in the preceding section, this section proposes suggestions on how to address these gaps and highlights the importance of addressing them in order to enhance the development of adaptive controllers for multiple service robots.

2.6.1 Mapping Scheme for pHRI

The proposed mapping for [pHRI](#) was essential to draw conclusions about human factors considerations in the literature. However, one cannot disregard the fact that the papers gathered by this review are limited to the ones that have addressed human factors, meaning that mapping for [pHRI](#) as a domain would require a broader range of studies. Thus, this mapping can be considered a starting point that is extensible to include more [pHRI](#) subcategories. For instance, as the field develops and more experiments are conducted, it is foreseen that the consistent subcategory in direct [pHRI](#) could be further broken down to have passive and active interaction, where robots may comply with human motion or not by applying forces against them (e.g. imposing the robot's task as prioritized with respect to the human intention/action). It is expected that such a category would impact robot perception, and usability [\[75\]](#).

Table 2.5: Quantification approaches employed within each measurable dimension, highlighting studies proposing quantification approaches. Studies using mathematical and machine learning models are categorized based on the input requirements of each model, including task parameters, physical data, and physiological signals.

HF	MD	Q	Machine Learning Model		PS	Mathematical Model			PD
			PD	PS		PD	TP	PS	
Cognitive Ergonomics	Mental Workload	[6, 50, 51, 32, 108, 66, 93, 188, 10]		[123, 111] *	[75] * [76]	[92] *	[175] *	[92] *	
	Stress	[156, 22]			[75] * [76, 38]	[92] *		[92] *	
Physical Ergonomics	Pain Sensitivity								[17] *
	Tactility	[25, 179, 157]							
						[175] * [48, 57, 39, 31, 99, 86, 124, 26, 27]			
	Physical Comfort	[5, 38, 167, 140, 70, 141]	[136, 135, 183] *		[38, 88]			[137] *	[117] * [188]
	Mechanical Trans- parency	[6, 39, 28, 46, 105]							[112]

* Studies that have proposed quantification approaches. HF: Human Factor Type, MD: Measurable Dimension, Q: Questionnaire, PS: Physiological Signals, PD: Physical Data, and TP: Task Parameter.

Table 2.5: *Cont.*

HF	MD	Q	Machine Learning Model		PS	Mathematical Model			PD
			PD	PS		PD	TP	PS	
Belief	Robot Perception	[159, 133, 152] * [126, 84, 69, 25, 156, 95, 56, 158, 50, 178, 160, 113, 73, 20, 11, 157, 170, 83, 68, 33, 120, 82, 43, 37, 77, 122, 41, 75, 76, 21, 132, 66, 110, 91, 148, 109, 61, 143, 105]			[14, 75] *				[75] * [60, 68]
		Trust	[60, 95, 6, 120, 43, 121]			[175] *	[175] *		[65, 177] *
		Perceived Intuition	[6, 51, 37, 75]		[75] *				[75] *
		Enjoyability	[50, 51]						
	Anxiety	[122, 41, 56]			[75] * [76]				
Affect	Emotional State	[69, 184, 49, 35, 178, 51, 33, 75, 91, 168, 20, 21, 170, 157]	[41]		[44, 75] * [96, 33, 181, 168]				[127] *
	Conveying Emotions	[4, 9]	[107] *						[69, 8] * [126]
	Receiving Emotions						[186] *		

* Studies that have proposed quantification approaches. HF: Human Factor Type, MD: Measurable Dimension, Q: Questionnaire, PS: Physiological Signals, PD: Physical Data, and TP: Task Parameter.

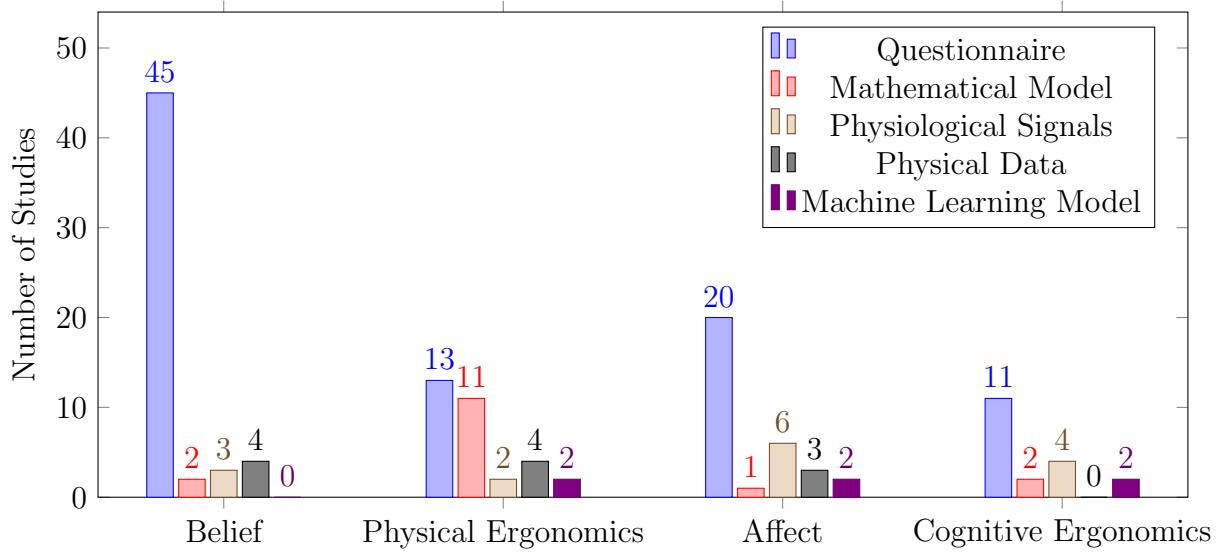


Figure 2.8: Number of studies relied on each assessment approach in each human factor type

Direct pHRI

The difference in the number of studies between direct and indirect pHRI domains reflects the research community’s greater focus on evaluating the human state during direct interaction conditions compared to indirect interaction conditions. This emphasis on direct pHRI may be attributed to the fact that many of the applications that have emerged are socially oriented, as evidenced by the greater number of studies on non-functional touch compared to functional touch. For example, handshaking, [171] [11] [14] [7] [127] [95] [82] [83] [113]; hugging [157] [56] [20] [170] [21]; clapping [50] [49] [51]; nursing [33] [32] [159] [22] [110]; and playing virtual games [123] [75] [76] [28] [141] have received significant attention. This trend is consistent with the idea that touch can serve important social and emotional functions in HRI, which in turn contributes to the realization of the Society 5.0 vision to a certain extent. However, it is important to note that the Society 5.0 vision encompasses assistive applications as well. Similar to the ongoing efforts to realize Society 5.0, the emergence of Industry 5.0, which emphasizes collaboration between humans and robots instead of full automation, is expected to drive the attention given to human factors in industrial settings. This includes the adoption of direct pHRI scenarios in industrial environments.

The identification of common studies between consistent contact and non-functional touch, such as handshaking and hugging, emphasizes the importance of investigating so-

cial behaviors in a realistic and broader context that includes more types of interactions. Conversely, the considerable overlap of studies between the categories of inconsistent contact and non-functional touch, especially in nursing scenarios, suggests a maturation of the field of **pHRI** in which researchers design experiments that encompass a wider range of interaction types beyond touch. Additionally, the studies that have investigated the purpose of direct **pHRI** (i.e., functional touch vs. non-functional touch) have mainly focused on nursing scenarios, indicating the relatively greater attention given to this area. However, the common studies detected between inconsistent contact and non-functional touch, which mainly comprise studies addressing social touches, may be due to the fact that touches occur spontaneously in human interactions and do not necessarily trigger or stop them if not maintained, unlike functional touches where many studies were exploratory and used the robot as a tool to follow certain paths.

The significant overlap between consistent contact and functional touch, and between inconsistent contact and non-functional touch suggest that they may be redundant, indicating a need to consider either the purpose or duration categories for future study classification in reviews. Further refinement of these classifications could enhance our understanding of direct **pHRI** as the field matures.

Our main objective is to bring robotic multiple service providers to life, and it is promising that most direct studies of **pHRI** focus on non-industrial applications. However, this also highlights a gap in current research for Industry 5.0. While non-industrial applications have explored social aspects of **pHRI**, such as handshaking, hugging, and playing, they have not adequately investigated physical assistance applications in daily life activities such as toileting, sit-and-stand, transferring, and others. One potential reason is that certifying robots for domestic environments is more challenging compared to industrial environments. Unlike controlled industrial settings, domestic environments are characterized by various uncertainties, making it harder to thoroughly examine potential applications in realistic settings. As a result, successfully transitioning from the lab to the actual domestic environment becomes more difficult. Without proper physical assistance, the potential benefits of social aspects are limited, and **pHRI** may not be useful in domestic or institutional environments. Therefore, it is important to investigate more physical assistive applications, while taking into consideration the human state, to fully realize the potential of robotic multiple service providers.

Indirect **pHRI**

Although human factors in handover scenarios have been explored in various contexts, there has been insufficient attention given to handovers in domestic settings. Handovers

in the industrial context are being thoroughly investigated, including the orientation of tools and the human body posture. However, the objective of comfortable and efficient handovers is likely similar in both industrial and domestic settings. This suggests that results from studies conducted in one setting could be applicable to the other.

However, most studies on atypical scenarios and assembly tasks are industrial-focused, highlighting the need to address domestic applications beyond those currently studied. Atypical scenarios in domestic settings, such as dressing, grooming, showering, or household chores, are complex and technically challenging. Similarly, assembly contexts in domestic settings, such as putting pillows into pillowcases or performing certain cleaning duties, could also pose technical challenges.

Studies on co-manipulation scenarios are limited, and the focus is mainly on reducing human joint overload during collaboration. Few studies have investigated mutual adaptation to achieve a common goal. While addressing joint overload is a logical first step in co-manipulation, mutual adaptation between humans and robots is also crucial in domestic scenarios where robots can assist with tasks such as furniture reorganization.

The scarcity of research in the indirect [pHRI](#) category underscores the insufficient attention that human factors researchers have paid to this field. Additionally, The lack of studies that addressed different applications simultaneously suggests that more research is needed in this area to gain a better understanding of the challenges and opportunities of indirect [pHRI](#), since many industrial settings may require multiple types of applications. Notably, domestic environments have received minimal to no attention in any of the sub-categories of indirect [pHRI](#). It could be argued that the term [pHRI](#) is more commonly used in industrial settings than in domestic contexts in the indirect [pHRI](#) domain.

2.6.2 Human Factor Classification

The holistic human factors model adapted to [pHRI](#) and proposed in Figure 2.1 is focused on our specific research questions and therefore differs from the original in some respects, i.e., in [36]. In particular, we did not include measurable dimensions related to "performance" as a human factor type under "ergonomics" since it was not a primary focus of our analysis. Based on the user experience definition, we found it appropriate to categorize all measurable dimensions related to physical and cognitive ergonomics under the "user experience" umbrella, whereas the original model depicts "user experience" overlapping with these human factor types but not entirely encompassing them. It is worth noting that for a more comprehensive understanding of the conceptual model, we recommend readers to refer to the original adapted [HRI](#) model.

Given the objectives of our review, we focused on factors that are affected *by* the interaction. Therefore, our review did not detect factors that appear out of the scope of the definition of user experience, such as workplace design and awareness, which are not impacted during an interaction [36].

We ensured that each measurable dimension we identified was classified under only one human factor type, as overlapping dimensions between types would have contradicted the holistic approach of the conceptual model we adopted. To achieve this, we based our classification on the definitions of usability, user experience, hedonomics, ergonomics, and their respective sub-categories.

Similarly, based on the definition of usability and the main objectives of this review, all of the measurable dimensions detected in usability belong to the hedonomics category of human factors. This is because the objective aspects of usability, which are more related to performance, were not a primary focus of the review.

Although defining each measurable dimension in detail was not our primary focus, we classified them based on the provided descriptions of human factor types and the usability definition. This approach allowed us to present an accurate and unbiased overview of the most relevant human factors that have been addressed in [pHRI](#) literature.

Our primary goal is to expedite the development of adaptive control frameworks for multiple service robots. We believe that investigating the most frequent human factor types evaluated together in a single experiment, along with exploring the relationships between measurable dimensions within each human factor type, will aid in this development by identifying the human factors that may have significant correlations for researchers to consider when designing adaptive control frameworks.

Since affect and belief address humans' emotions from different perspectives, it is not surprising that most of the studies that have evaluated more than one human factor type belonged to belief and affect categories. Similarly, cognitive ergonomics and belief come next in the most common number of studies between human factor types, since both evaluate cognitive aspects but from different perspectives. In contrast, no studies have evaluated humans' emotions during an assessment of their physical ergonomics. Therefore, it is anticipated that there could be an unrevealed relationship between the three human factor types and their measurable dimensions that is worth investigating. On a broader view, unrevealed relationships between measurable dimensions within a human factor type are very likely to exist. However, a lack of definitions may not support conducting such a study, unlike correlating human factor types to each other, which can be perceived as more feasible.

Direct [pHRI](#) studies addressed a wider variety of human factor types in comparison

with indirect [pHRI](#) studies since many more papers have emerged in the former than the latter. As expected, affect is the most addressed human factor type among the non-functional touch studies, and accordingly, the inconsistent studies, given the big number of common studies among both categories. Along the same lines, physical ergonomics is the least addressed human factor type among them since those type of studies is more socially directed. It was foreseeable that belief would be highly addressed among most of the studies since [HRI](#) literature has pointed out that dimensions such as trust were the most frequently addressed in comparison with any other dimensions. In contrast, cognitive ergonomics was seldomly adopted.

As previously mentioned, the majority of indirect [pHRI](#) applications are focused on industrial scenarios. Consequently, it is not surprising that physical ergonomics is the most frequently addressed category, with limited attention given to affect. In fact, affect was hardly mentioned in any of the indirect categories. Belief, however, which takes emotions into account, received more attention than affect. This could be due to its evaluation being influenced by cognitive aspects.

To fully realize the potential of [pHRI](#) in domestic scenarios, it is essential to address various gaps in the literature. In direct [pHRI](#) scenarios, the robotics research community must overcome the challenges of conducting physical assistance experiments, such as toileting, and sit-and-stand, and assess those applications from a human factors perspective. While social interaction experiments like hugging, handshaking, and playing are important, it is crucial to combine both physical interaction types to understand their influence on user perception during physical assistance. Additionally, applications that may be considered industrial or non-industrial, such as handovers and co-manipulation, may benefit from addressing hedonomics and cognitive ergonomics, besides physical ergonomics. Moreover, atypical and assembly applications such as eating assistance, grooming, showering, and household chores require attention to various human factors. Addressing these literature gaps will enhance the overall [HRI](#) experience and increase the usefulness of [pHRI](#) in domestic settings.

We believe that researchers aiming to address the identified literature gaps can find valuable insights in Table 2.4. This table outlines the measurable dimensions investigated for each [pHRI](#) category. Researchers can utilize this resource to:

1. Conduct studies in physical assistive scenarios.
2. Combine various physical interactions.
3. Explore the generalizability of proposed quantification approaches by applying them in different scenarios.

4. Examine potential correlations between measurable dimensions within a human factor type.
5. Uncover relationships between different human factor types.
6. Investigate human factor types not explored in specific [pHRI](#) scenarios.

By leveraging the quantification approaches provided in Table 2.5, researchers can investigate the impact of employed physical interactions on the significant measurable dimensions they identify, based on their available resources and research objectives.

It is worth mentioning that one may argue that certain literature gaps, such as assessing mental workload or stress during a handshake, or emotional states during co-manipulation, might not be as immediately useful as evaluating physical ergonomics, which is valid. However, if our goal is to deploy robots for various tasks, i.e., multiple service robots, we must ensure that these robots can understand the overall human state during different applications. They shouldn't be limited to performing single actions, like handshakes, hugs, or co-manipulation, but rather be versatile enough to carry out a wide range of tasks.

2.6.3 Human Factors Quantification

Due to the lack of consensus on human factors terminology and definitions in the domain of [HRI](#), the large number of experiments conducted to understand the impact of human factors is not expected to yield efficient and uniform progress in the field. In other words, without reliable and agreed-upon quantification approaches, the results of experiments on the impact of human factors may not be reliable enough to facilitate the development of the field. However, a promising aspect is that a significant number of studies proposed quantification approaches and validated them, demonstrating the research community's awareness of the importance of such validation. The development of standardized and reliable quantification methods has the potential to facilitate the detection of progress in the field of [pHRI](#), thus accelerating its advancement.

It is clear that the human factor types addressed in Figure 2.2 follow a trend similar to the measurable dimensions addressed in Figure 2.3. However, the human factor types are comprehensive, reflect the research direction in a broader view and give all measurable dimensions, within a human factor type, the same level of importance despite the terminology used. Thus, adopting the conceptual model is expected to yield efficient uniform progress in the field.

Figures 2.6 and 2.7 demonstrate that subjective approaches are the primary evaluation method used across all human factor types. Questionnaires are commonly used due to

their simplicity and fast administration, whereas other assessment approaches may require specialized equipment and complex data processing. These findings confirm our hypothesis for **RQ3**, which was based on the results obtained from the [HRI](#) field. However, questionnaires should be considered a preliminary assessment approach that provides an initial understanding of human states or validates other objective approaches. Otherwise, the development of adaptive control frameworks for human states is unlikely to become a reality in the near future. Hedonomics relies primarily on questionnaires, whereas ergonomics has investigated objective approaches more frequently than subjective ones. Physical ergonomics parameters are quantifiable, which explains why objective measures have been used in physical ergonomics studies more than in cognitive ergonomics. Future efforts should focus on developing objective measures for the other three human factor types to enable the implementation of real-time adaptive control frameworks in the future. Nevertheless, developing real-time objective quantification approaches for [HRI](#), including its subcategory [pHRI](#), is challenging, which explains why these approaches are rarely addressed in these fields.

Quantifying human factors based on measurable data involved various approaches, including identifying correlations, and causal relationships; developing machine learning models; and formulating mathematical models. While identifying correlations between measurable dimensions and physical or physiological data can aid in quantifying certain human factors, it is essential to acknowledge that these correlations might not hold true in every other application. For validation and confirmation of such correlations, a multitude of empirical studies must be conducted across diverse scenarios and with a substantial number of participants. Only through rigorous examination can correlations be deemed reliable across various applications.

For instance, consider the significant correlations found between participants' hand temperature and their positive attitudes towards robots in handshake scenarios [14]. The question arises: will the same significance persist in another scenario where participants exercise with a robot by pushing against its arm or utilize a robot to help them maintain their balance in certain scenarios? This raises the issue of generalizability, which presents an enduring research challenge that remains unresolved.

Conversely, conducting experiments to establish causal relationships, which are often regarded as more reliable than correlations, can be a costly endeavor, prompting researchers to seek more cost-effective alternatives. Machine learning models hold promise in this regard, although they come with the limitation of requiring extensive training on diverse datasets to produce dependable results, and collect sufficiently large datasets in [pHRI](#) is extremely challenging.

Mathematical models, on the other hand, offer generalizability but lack flexibility. They

are better suited for scenarios where the relationships between variables are well-defined and unchanging. In contrast, machine learning models exhibit greater flexibility, allowing them to adapt to varying patterns and complex data. However, this flexibility can also hinder their generalizability, necessitating careful consideration of the application domain and dataset.

In summary, achieving robust quantification approaches of human factors based on measurable data demands extensive empirical studies and careful examination across multiple scenarios. Researchers should weigh the advantages of machine learning models, which offer flexibility but require substantial training data, against the benefits of mathematical models, which provide generalizability but may lack adaptability in more complex situations. Additionally, it is important to view correlations and questionnaires as valuable approaches and intermediate stages that can aid in creating robust adaptive controllers for robots, but not the ultimate approaches for quantifying human factors. By keeping these points in mind, researchers can deepen their understanding of human factors and effectively utilize suitable quantification approaches to predict human states.

Researchers seeking to address the identified gaps in the literature can benefit from Table 2.5, which provides valuable insights into quantification approaches for each measurable dimension. Depending on their resources and objectives, researchers have the option to enhance existing quantification methods as outlined in mathematical modeling, machine learning models, and correlations. These choices can be guided by the advantages and limitations mentioned.

2.7 Limitations of the review

This section is dedicated to demonstrating the validity of this review by discussing its limitations. It should be noted that limitations and future work of the field are discussed within Section 2.6, Discussion. In order to evaluate the validity of this review, 3 validity factors proposed by [138] can be applied: theoretical validity, interpretive validity, and descriptive validity.

A common limitation in literature reviews is the possibility of existing studies that are related to the main topic but could not be detected due to unfortunate choices of keywords either from the review’s authors or the related studies. However, the adopted guidelines, which included the test-set step, could make us argue that this type of threat is eliminated to a decent extent. Petersen et al. [138] have referred to this validity as theoretical validity which they defined as “our ability to be able to capture what we intend

to capture”. Therefore, theoretical validity includes data extraction validation. In order to restrain data extraction bias, full-text checkout over all the gathered papers was performed iterative, until a relatively proper lumping of measurable dimensions is achieved.

Another validity factor is interpretive validity, which “is achieved when the conclusions drawn are reasonable given the data” [138]. To limit interpretive validity threats, HRI experts got involved during each step of conducting this review, including the interpretation of the extracted data. However, one can argue that broad dimensions, such as robot perception, emotions, and physical comfort are misinterpreted as they can be divided into many smaller dimensions. For instance, robot perception can be divided into attitudes, acceptance, perceived performance, etc. However, according to the HRI conceptual model adopted, measurable dimensions that do not intersect between human factor types cannot bias our results as previously explained in Subsection 2.5.2. Nevertheless, breaking down factors can provide a clearer understanding of the measurable dimensions addressed within a human factors type. Therefore, it is highly encouraged for future reviewers to consider this approach. Furthermore, it can be argued that one common human factor, such as perceived safety, did not emerge as a measurable dimension within any human factor type, which might be regarded as another misinterpretation. However, perceived safety is a broad term that is shared between different human factor types as it includes dimensions such as stress, anxiety, trust, fear, and psychological comfort[146]. Therefore, considering it as a measurable dimension is not reasonable.

To maintain descriptive validity, which “is the extent to which observations are described accurately and objectively”, the adapted model of human factors in HRI is considered. The conceptual model provides clear outlines for usability, user experience, and each of the human factors types, while still allowing for accurate classification of measurable dimensions within each category. However, it can be noticed that performance as a human factor type within ergonomics is discarded, which can rise a descriptive validity threat by arguing that ergonomics is not accurately conducted throughout this review. As discussed in Subsection 2.5.2, performance, as defined by Coronado et al. [36], includes the evaluation of the performance of the robot, the user, and the overall system. However, since our main objective is to quantify the human state, these performance evaluations are not directly relevant to our review. Especially that perceived performance, which contributes to the human state, is included in robot perception, as shown in Table 2.3. As this is a human-centered review, research has shown that in domestic contexts, humans tend to prefer robots that exhibit human-like characteristics over high-performance robots [66]. Therefore, since our primary objective is to quantify the human state, discarding performance evaluations do not negatively impact the main focus of this review. Therefore, we are confident that the conclusions of this review can serve as a reliable foundation for

future research in this field.

2.8 Conclusion

For multiple service robots to coexist with humans, they must meet users' physical and social needs. Socially, robots must be able to predict and quantify human states to interact appropriately. Physically, manipulation is a key capability. This chapter focused on studies of robots with manipulation abilities and their impact on quantifiable human factors in [pHRI](#). The goal is to advance the development of [pHRI](#) control frameworks that can predict human states and make decisions accordingly.

To achieve the objective of this review, we identified the most common quantifiable human factors in [pHRI](#), noted the factors typically addressed together and determined the most commonly used assessment approaches. We also collected and classified proposed assessment approaches for each human factor and created an initial map of the common contexts and applications in [pHRI](#). Our aim was to provide an unbiased overview of the field, identify research gaps, and facilitate the search for adaptive control framework advancements.

Most studies in the direct [pHRI](#) category focus on social behaviours, with belief being the most commonly addressed human factor type. Task collaboration is moderately investigated, while physical assistance is rarely studied. In contrast, indirect [pHRI](#) studies often involve industrial settings, with physical ergonomics being the most frequently investigated human factor. More research is needed on human factors in direct and indirect *physical assistance applications*, including studies that combine physical social behaviours (e.g., comforting touch) with physical assistance tasks (e.g., sit-to-stand). This will enable robots to exhibit appropriate social skills while providing physical assistance. Expectedly, the predominant approach in most studies involves the use of questionnaires as the main method of quantification. However, it is worth noting a recent trend that seeks to address quantification approaches based on measurable data.

To advance service robots and Society 5.0, future reviews should include a wider range of [HRI](#) scenarios beyond physical collaboration. These reviews should prioritize studies that use measurable data to quantify human factors, as well as those that explore correlations between demographic and personal information and human factors. By identifying relationships between measurable dimensions of each human factor, we can create more robust adaptive control frameworks for various [pHRI](#) scenarios.

2.9 Aligning with Thesis Objective

This chapter has been instrumental in addressing the two fundamental questions: 'What is human mental state?' and 'How can we quantify human mental state?' Through this exploration, we identified cognitive ergonomics and belief as the two human factor types associated with cognitive aspects that best represent the human mental state in our study [1]. These dimensions encompass robot perception for belief and mental workload and stress for cognitive ergonomics.

To accurately measure each dimension, a combination of objective and subjective approaches is essential. Table 2.5 facilitated the identification of appropriate tools for each approach. For belief dimensions, objective quantification includes measures such as skin temperature [14]. While RoSAS questionnaire, supplemented by perceived safety dimension from the Godspeed questionnaire [16] can serve as the subjective approach [29]. In cognitive ergonomics, objective quantification methods include techniques such as Photo-plethysmography (PPG) [75], monitoring skin temperature of human body peripherals [85], and assessing task performance [24]. However, when it comes to subjective assessments, we rely on standardized questionnaires from domains beyond robotics due to the absence of an accurate subjective questionnaire tailored to our needs. Notably, the NASA Task Load Index (NASA-TLX) [71] questionnaire, while widely recognized, has shown major limitations such as multi-dimensionality and the inability to accurately differentiate between mental load and mental effort [23], indicating its inadequacy. The difference between mental load and mental effort is represented in Section 5.4.3, Measures, in Chapter 5.

Evaluation of subjective approaches for both cognitive load and belief is crucial. In this regard, we aim to assess the subjective quantification approach for robot perception as a first step. Despite being the primary focus for pHRI researchers, robot perception has lacked validated quantification approaches, with a minimal exploration of reliable objective measures. Through an independent user study outlined in the subsequent chapter, we seek to evaluate common subjective quantification approaches for robot perception. Only by reliability assessment of these subjective approaches can we discern well-grounded correlations between subjective and objective quantification methods, such as skin temperature. However, the evaluation of subjective approaches for cognitive ergonomics remains an unaddressed gap in pHRI literature, warranting further investigation.

The subsequent chapter, i.e., Chapter 3, is dedicated to elucidating the user study conducted to gather the essential data for the assessment of the subjective quantification approach for robot perception. Chapter 4 presents the analysis and results of this approach's assessment in direct pHRI scenarios. Furthermore, Chapter 5 explores correlations between subjective and objective approaches to belief and cognitive ergonomics.

Chapter 3

The Impact of Robot’s Facial Emotional Expressions on Light Physical Exercise

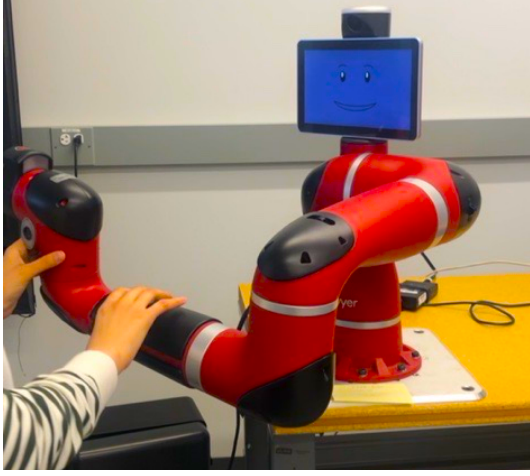
3.1 Introduction

As previously mentioned, [pHRI](#) plays a crucial role in supporting independent living for older adults. A potential enhancement for successful aging is introducing robots that can provide physical assistance with essential daily activities and promote light physical exercises. It has been proven that light physical exercise can be beneficial for older adults to maintain their cognitive and physical well-being [52]. Especially during the COVID-19 pandemic, the use of robots has significantly increased, not only in environments such as hospitals and factories but also at home [58, 12], with recent studies showing that people’s perception of robots has significantly changed towards more positive directions in terms of acceptability [59]. The shift in demographic structure, the adoption of robotic technologies, and the evolution in people’s mindsets concerning these robots drew our attention to exploring what could contribute to a successful [pHRI](#) [101], especially considering that most of the existing works in this direction are using social robots that do not necessarily physically interact with human users.

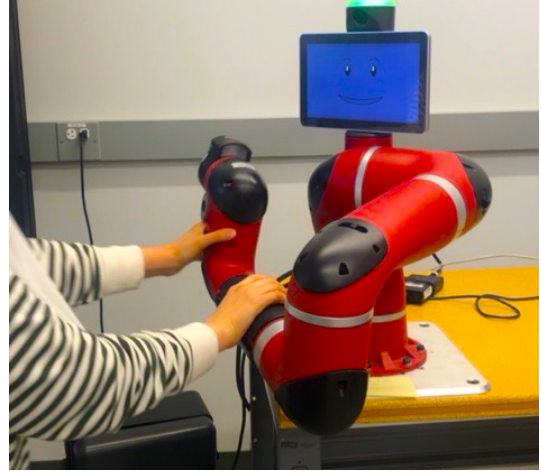
Recent studies have demonstrated that, by bridging [Social Human-Robot Interaction \(sHRI\)](#) and [pHRI](#), robots can physically interact with humans while also being socially acceptable [52]. One approach that has proven its effectiveness in enhancing robot perception and engagement is endowing robots with the ability to exhibit facial emotional expressions

[173]. This has motivated us to explore the potential of facial emotional expressions in pHRI scenarios.

In our study, we adopt a light physical exercise scenario, as one of the potential applications for physically interactive robots in domestic environments, to investigate the impact of facial emotional expressions on users' perception of the robot and their level of engagement. We utilized Sawyer, a collaborative robotic arm developed by Rethink Robotics, for this purpose. Sawyer is equipped with a flat panel screen that allows us to display various facial expressions, as shown in Figure 3.1.



(a) Before performing a push, the robot is in its initial configuration



(b) After a push is performed, the robot head-light blinks

Figure 3.1: Participant performing a physical pushing exercise with Sawyer displaying a facial expression on the flat screen panel.

The primary motivation for this user study was to enhance healthy aging, which initially involved 27 older adults aged 65 years and above. However, to ensure robustness and reliability of the subjective tool assessment across various age groups, and to explore notable differences observed in preliminary experiments with young, healthy participants and the older adult population, we aimed to gather a larger, diverse sample size. Our objective is to investigate potential variations between age groups concerning facial expressions and physical exercises with a robot. Literature suggests that utilizing robots for motivating exercise extends beyond older adults to other age groups [18, 64]. We anticipate that our results will be helpful for the research community toward the exploration of effective social skills for physically interactive robots for both age groups. Chapter 4 presents a detailed analysis of the dataset acquired to evaluate the reliability of our subjective assessment tool,

while this chapter focuses on elucidating the user study and its outcomes independently.

This chapter is structured as follows: Sec. 3.2 provides an overview of existing research on facial emotional expressions and their use in HRI, as well as an overview of the effect of age on attitude towards robotics technology; in Sec 3.3, the research questions guiding the study are presented; Sec. 3.4 details the experiment design, participants, and measures taken for evaluation purposes; Sec. 3.5 illustrates the obtained numerical and qualitative results; Sec. 3.6 discusses the obtained results with respect to the outlined research questions; Sec. 3.7 outlines the limitations of the presented study, providing insights for possible future improvements and developments; Sec. 3.8 concludes with a brief summary and take-aways.

3.2 Related works

Our research is founded upon two primary domains: [sHRI](#) and [pHRI](#), which serve as the pillars of our investigation. Remarkable efforts were made to explore social-physical robots in various contexts such as hugging [21], touching in social and nursing scenarios [110], handshaking [95], and playing games [49]. However, only limited efforts involved the investigation of both domains with robots that possess high dexterity and manipulation capabilities [1] which are essential qualities for robots to efficiently engage in physical interactions. Even fewer studies have devoted their efforts to investigating facial emotional expressions, for those types of robots [49]. Furthermore, most studies have focused on either a general population (typically younger adults) or older adults separately, without a direct comparison between the two age groups, thus not providing sufficient insight into possible differences and commonalities between age groups.

3.2.1 Facial expressions in HRI

Significant efforts have been dedicated to endowing social robots with the capability to display facial emotional expressions. Lin et al. [97] investigated the association of colors with robot facial emotions for the social robot Zenbo. According to their study, participants associated the color blue with both positive and negative emotions conveyed by the robot. However, the researchers found that participants were only able to correctly identify the robot’s facial emotions about half of the time. Notably, positive facial emotions were identified more accurately than negative ones. This contradicts the results found by Fitter and Kuchenbecker [53], who designed facial emotional expressions, especially for the Baxter robot. Their study showed that all participants across different cultures and despite the

robot’s facial color were able to correctly map the six basic emotions on the valence-arousal model [147]. It is important to note that while both studies evaluated rendered robot facial emotional expressions across cultures, Ekman et al.’s six basic emotions [42] have been proven to be universally recognized, unlike the diverse set used by Lin et al. Additionally, it is worth considering that the set of faces used by Fitter and Kuchenbecker might have had more abstract facial features than the corresponding set used in Lin et al.’s study. Moreover, one can argue that the participants’ ability to differentiate between facial emotions has been influenced by the disparity in the number of emotions they were required to identify in each study.

Some existing studies have relied on facial expressions to enhance the user’s engagement with robots that possess high dexterity and manipulation capabilities in various pHRI scenarios, such as physical exercise [52], clapping/gaming, and teaching [1]. However, they did not investigate the impact of the robot’s facial expressions on the interaction, unlike Tsalamlal et al. [171] and Fitter et al. [49]. Tsalamlal et al. [171] investigated how participants combine facial expressions and handshakes to assess the perceived emotions in robots. The findings indicated that participants assigned greater significance to facial expressions when evaluating Valence. Fitter et al. [49] assessed how participants’ emotions were affected by a robot’s responsive facial expressions compared to an unresponsive robot’s facial expressions during a hand-clapping game. Participants perceived the interactive face as more pleasant, energetic, and less robotic than the unresponsive one.

However, none of the previously mentioned studies have investigated participants’ perceptions of the robot’s characteristics, such as its perceived intelligence. Furthermore, the influence of these facial expressions on users’ performance remains unexplored. To address these gaps, our study seeks to investigate participants’ perceptions of a physically interacting robot displaying facial expressions and examine the potential impact of these expressions on their performance during physical interactions.

3.2.2 Age groups differences

Aside from common societal stereotypes, research has proven that there are identifiable differences between age groups when it comes to the adoption and perception of technologies, which can be influenced by various factors, including perceived usefulness, ease of use, experience, and self-efficacy.

While older adults’ adoption of technology has historically been lower than younger adults, it is steadily increasing. Understanding the benefits of technology, such as improvements in cognitive, social, and emotional well-being, is crucial. Positive experiences

and the opportunity to gain technological self-efficacy are significant predictors of technology adoption among older adults. The gap in digital technology adoption between older adults and younger generations has narrowed, with notable increases in smartphone ownership, social media use, and internet use among the older population. However, differences remain, such as in the frequency of internet use and ownership of certain technologies like tablets. [114].

The attitude towards technology is reflected in robots as well, being a new type of technology that has slowly entered the market only in the past couple of decades. Research in this area has analyzed aspects like comfort levels with robot assistance, perceived utility and trustworthiness of robots, and the potential for robots to provide social or physical support [103]. Studies have shown that older adults see robots more as assistive devices that can help them maintain their independence or provide companionship [102, 161], while younger adults focus more on the efficiency and novelty aspects of robotic technology [13, 45]. As with other technologies such as smartphones and tablets, these perceptions may also be influenced by familiarity with technology and perceived ease of use.

However, very few studies have directly compared the use of a robot platform under the same conditions to evaluate the possible differences between the two age groups, and even fewer have evaluated such differences in the domain of pHRI. Fitter et al. [52] obtained results showing that older adults had an overall more positive experience interacting with a robot during physical exercises than younger adults, but their focus was not reflected on the impact of facial expressions. Furthermore, recent studies have shown that people’s attitudes, both for young and older adults, have changed significantly during the COVID-19 pandemic [59]. To bridge gaps in existing studies, our study extends the investigation of facial expressions and their impact on physical exercises with a robot to a comparison between an older and a younger age group.

3.3 Research Questions

Our study is designed to build upon the insights gained from the existing literature presented in the previous section. By rigorously exploring the effects of facial emotional expressions on both older and younger adults engaged in pHRI applications. We first aim to investigate the two age groups separately and then perform a comparison between the two. With older adults, in this study, we refer to a population with ages greater than 65 years old, and with young adults, we refer to a population of age between 18 and 30 years old.

The following first set of research questions were formulated to lead this study:

- **RQ1.1:** Will facial emotional expressions impact an older adult’s level of engagement and perception of a robot in a pHRI scenario?
- **RQ1.2:** Will facial emotional expressions impact a younger adult’s level of engagement and perception of a robot in a pHRI scenario?

The answer to the first set of research questions will help us understand the importance of relying on facial expressions as a means of communication during pHRI with both older and younger adults. These findings will guide the research community towards investing further efforts in the development of facial emotional expressions for successful pHRI. Alternatively, they may prompt exploration of other social behaviors that could be better suited for typical pHRI scenarios.

To comprehend the impact of facial expression responsiveness, as well as the mere presence of a robot’s face regardless of its responsiveness, on participants’ perceptions of the robot’s characteristics and performance, we pose our second set of research questions:

- **RQ2.1:** Does a responsive robot’s facial emotional expressions impact an older adult’s level of engagement and perception of a robot, compared to an unresponsive robot, in a pHRI scenario?
- **RQ2.2:** Does a responsive robot’s facial emotional expressions impact a younger adult’s level of engagement and perception of a robot, compared to an unresponsive robot, in a pHRI scenario?

To understand possible differences and common factors between older and younger adult populations with respect to the above two sets of research questions, we formulated the following third set of research questions:

- **RQ3.1:** Are there significant differences between older and younger adults on the impact of facial emotional expressions on the level of engagement and perception of a robot in a pHRI scenario?
- **RQ3.2:** Are there significant differences between older and younger adults on the impact of a responsive robot’s facial emotional expressions on the level of engagement and perception of a robot, compared to an unresponsive robot, in a pHRI scenario?

3.4 Methodology

To answer the proposed research questions, we designed a user-study. The details of the study, including the design of the experiment, the participant pools, the procedure of the experiment, and the utilized metrics are presented in this section.

3.4.1 Experiment design

To answer the proposed research questions, an in-laboratory user study, in the form of a light physical exercise game with a collaborative robot arm with a flat-screen display, was conducted at the Active & Interactive Laboratory at the University of Waterloo in Ontario, Canada. The objective of the exercise is to perform the highest number of pushes possible, according to each participant’s comfortable pace, against the robot’s 4th joint as shown in Figure 3.1. Further details of the game design are provided in Section 3.4.1.

A between-subject study design is considered and Sawyer from Rethink Robotics is used. Sawyer is a 7-degree-of-freedom torque-controlled manipulator equipped with a flat-panel screen and headlight, as shown in Figure 3.1.

During participants exercising with the robot, 3 conditions of facial emotional expressions were considered:

1. **Inactive Facial Expression:** The robot displays its default screen, which features the Sawyer SDK logo from Rethink Robotics [145], as depicted in 3.2.
2. **Unresponsive Facial Expression:** The robot showcases a happy face (as depicted in Figure 3.3b).
3. **Responsive Facial Expression:** The robot exhibits varying facial emotional expressions in response to the user’s performance, with three variations as depicted in 3.3.

In this chapter, we will also refer to the unresponsive and responsive conditions together as the active conditions. In the responsive condition, the robot shows a neutral face (shown in Figure 3.3a) at the beginning of the interaction. After the participant performs a set of successful pushes, the robot shows a happy face (shown in Figure 3.3b) and after another set of successful pushes, the robot shows a surprised face (shown in Figure 3.3c).

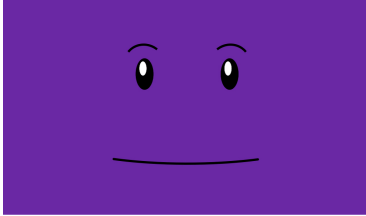
We opted to utilize Fitter and Kuchenbecker’s established facial emotional expression set [53] due to its cross-cultural evaluation, a critical factor for conducting experiments

in a Canadian societal context, as in our case. Fitter et al. [53] found that participants from the USA and India successfully identified their proposed set of facial emotional expressions. Moreover, the study, which took place online, exclusively presented participants with Baxter’s head. It is worth noting that Baxter’s head is almost identical to Sawyer’s head, as both robots are products of Rethink Robotics.

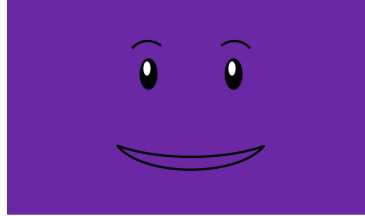
We considered using the safest rated facial emotions, according to Fitter and Kuchenbecker’s results, as safety is the most crucial human factor in HRI [36]. Therefore, we decided to use the neutral, happy, and surprised faces for the responsive condition and the happy face, which is rated the safest among all faces, for the unresponsive condition. While the red and the purple colors were rated as the most energetic face colors, we chose to use purple for all conditions as red was, also, rated the least pleasant. Similarly, we aimed at using an arousing color as it contributes significantly to promoting the interaction. It should be noted that Sawyer’s face color in Figure 3.1 is purple, but it may appear blue due to the camera effect.



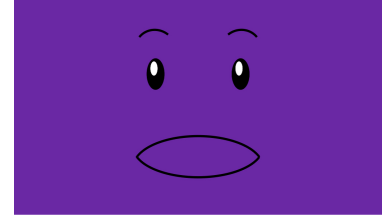
Figure 3.2: Sawyer SDK logo displayed as Inactive Facial Expression



(a) Neutral Face



(b) Happy Face



(c) Surprised Face

Figure 3.3: Sawyer facial emotional expressions [53], 3.3b is used as the facial expression for the Unresponsive Facial Expression condition.

Physical Exercise Design

The exercise starts with Sawyer at its initial joint configuration, as shown in Figure 3.1a. The participants stood upright and faced Sawyer’s head display. They are asked to perform

the highest number of pushes possible according to their comfortable pace in 1 minute against Sawyer’s 4th joint (the ”elbow”) as shown in Figure 3.1. Further details about the instructions provided for the participants are indicated in Section 3.4.3. A push is only counted if a participant was able to push the joint to make an angle offset greater than a pre-defined threshold. Each time a push is counted, Sawyer’s green headlight blinks as an indicator for the participant. Figure 3.1b shows the robot state, note the headlight, when a push is counted. The actual exercise performed by a participant can be found in the video attachment.

Robot Control

To facilitate the physical exercise with Sawyer in a fully autonomous mode, the robot control was implemented using a joint impedance control scheme [131]. The joints’ initial configuration and stiffness are kept constant across all conditions, whereas Sawyer’s facial expressions are adjusted according to the condition being examined.

In the responsive condition, to keep a consistent interaction experience with Sawyer across all participants and account for differences in participants’ physical capabilities, each participant performed a trial session to determine their capability of pushing Sawyer’s joints, i.e., baseline. Further details on how the trial session is conducted are in Section 3.4.3. Thus, we were able to predict each participant’s total number of pushes during the actual session. Accordingly, we implemented our code to show a happy face after 25% of the baseline, and a surprised face after 75% of the baseline. Hence, each participant in the responsive conditions can experience all the facial expression alterations around the same phase in the experiment despite the expected diversity in participants’ physical capabilities.

3.4.2 Participants

For the **older adults** population, twenty-seven participants were recruited in our study (17 female (F); 10 male (M), all older adults) from the University of Waterloo Research in Aging Participant Pool (WRAP), between the ages of 66 and 88 years ($M = 76.52$, $SD = 6.12$). Efforts were made to ensure age and gender balance within each condition. Additional details regarding the sample are provided in Table 3.1.

For the **younger adult** population, thirty-four participants were recruited (19 female (F); 15 male (M)) from the University of Waterloo main campus via recruitment posters posted mainly outside of the Faculty of Engineering to avoid familiarity with robots among participants, between the ages of 18 and 39 years old ($M = 23.4$, $SD = 4.48$). Efforts were

Table 3.1: Physical Exercise pHRI experiment sample description for Older Adults

	Inactive	Unresponsive	Responsive
No. of Participants	9	9	9
Female Participants	5	6	6
Male Participants	4	3	3
Age in years	76.78 ± 7.31	77.44 ± 4.27	75.33 ± 6.18

Table 3.2: Physical Exercise pHRI experiment sample description for Younger Adults

	Inactive	Unresponsive	Responsive
No. of Participants	12	11	11
Female Participants	7	6	6
Male Participants	5	5	5
Age in years	23.17 ± 5.3	22.9 ± 2.5	24.3 ± 4.9

made to ensure age and gender balance within each condition. Additional details regarding the sample are provided in Table 3.2.

For the older adults population, among the recruited participants for the experiment, the majority were right-handed. However, in the inactive condition, there were four exceptions: two left-handed individuals and two who identified as ambidextrous. Additionally, each of the two active conditions included one left-handed participant. Ethnically, the majority of participants identified as Caucasian, with two participants identifying as Southeast Asian. Notably, two participants of South Asian descent were assigned to each of the active conditions. Furthermore, all participants demonstrated good eyesight as confirmed by a brief eye test.

Initial survey results indicated that all participants (both older and younger adults) displayed normal levels of depression, stress, and anxiety according to the Depression Anxiety Stress Scale 21 (DASS-21) [125]. Notably, depression levels were evaluated due to their established influence on activity motivation, a practice observed in prior similar experiments [52]. None of the participants had prior exposure to the Sawyer robot, as confirmed by a 5-point Likert scale. While some participants had encountered other robots before, over 50% of the assigned participants had no previous robotics experience across all conditions. During the trial session, participants demonstrated closely matched physical

capabilities. Additionally, none of the recruited participants had upper or lower limb motion disabilities.

All our experiments received ethical approval from the University of Waterloo Human Research Ethics Board (protocol N. 45340) at the University of Waterloo, Ontario, Canada. Before the experiment, participants received proper information and gave informed consent to participate in the study.

3.4.3 Procedure

Each participant visited the laboratory and dedicated 20 to 30 minutes to complete the study. The procedure is as follows:

1. After obtaining participants' informed consent, they provided demographic information including age, gender, ethnicity, profession, and handedness. Following this, participants watched an instructional video ¹ on how to exercise with the robot, without indicating the robot's capability of facial expressions, i.e., its screen is not shown. The decision to withhold information about the robot's facial expression capabilities prior to the experiment was intentional, aiming to prevent the formation of unrealistic expectations.
2. A short eye test was then administered to ensure participants' ability to see the robot's screen clearly, this test was mainly conceived for older adults as eyesight tends to deteriorate with older age, however, to maintain identical conditions for the two populations for the purpose of comparison, the same test was administered also to the younger adults.
3. At the beginning of the setup, Sawyer's screen was turned away from the participant to prevent visibility of the screen. Participants were given the opportunity to perform 2 to 3 pushes to become accustomed to the robot's stiffness and determine their preferred distance from it. The experimenter ensured that participants performed the exercise correctly by ensuring they understood how the pushes were being counted.
4. A trial session lasting 10 seconds was conducted with Sawyer's screen still turned away. The purpose was to gauge each participant's physical capability. Upon successful completion of the trial, the robot rotated its screen to face the participant and proceeded to execute one of the three conditions detailed in Section 3.4.1. Participants conducted the trial session with the robot in the same state as depicted in the instructional video (Sawyer's screen not facing the participant).

¹[Instructional video](#)

5. Participants performed the actual session for a duration of 1 minute.
6. Following the task, participants completed robot perception and engagement questionnaires, detailed in Section 3.4.4. Subsequently, a debriefing session was held to address any questions or concerns. As a token of appreciation for their time, each participant received remuneration.

3.4.4 Measures

To evaluate robot perception, we employed the RoSAS [29]. RoSAS measures participants' perceived competence, warmth, and discomfort on a 9-point scale. Additionally, perceived safety was evaluated using the corresponding subscale from the Godspeed questionnaire [16], which employs a 5-point scale. Perceived trust was assessed with a single-item questionnaire employing a 5-point Likert Scale. Participants indicated their level of trust by responding to the statement 'I trust the robot,' where 1 denoted 'strongly agree' and 5 denoted 'strongly disagree'. Items within each sub-scale are randomized.

For engagement assessment, we employed both objective and subjective evaluation methods. Participants indicated their level of engagement by responding to the statement 'I felt engaged with the robot during exercising' using a 5-point Likert Scale, where 1 represented 'strongly agree' and 5 represented 'strongly disagree'. The objective assessment (E_{obj}) involved determining the ratio between the actual number of pushes during the session (P_{actual}) and the expected number of pushes ($P_{expected}$) calculated from the trial session.

$$E_{obj} = \frac{P_{actual}}{P_{expected}}$$

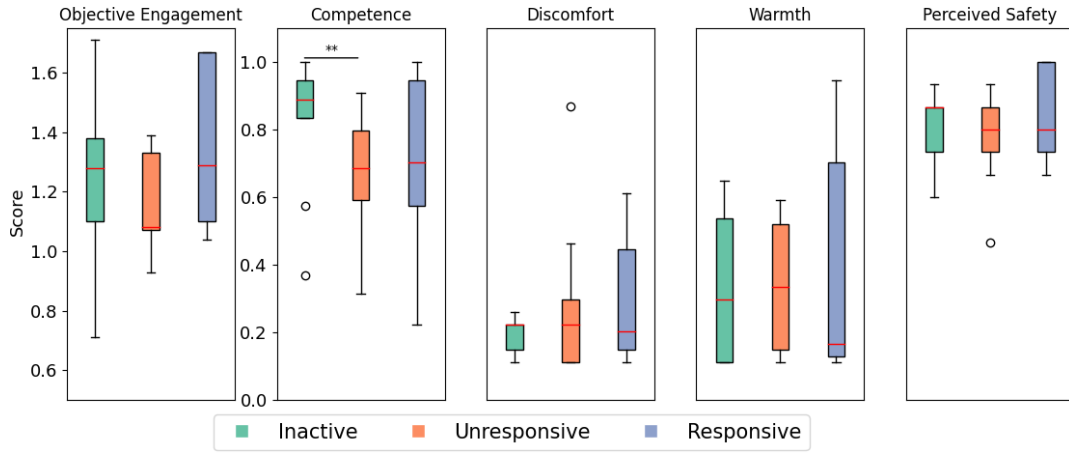
Furthermore, we sought to understand the reasons behind the participants' responses to our quantitative measures by relying on open-ended questions. These questions included: "Did you enjoy the exercising session? Why or why not?", "What do you think about Sawyer as an exercising partner?", "What stood out to you the most about interacting with the robot?", "Do you think the robot can have more features that would make it more interesting? Suggest features.", and for the responsive condition, "Did you observe any changes in the robot's facial expressions? If so, how would you describe the changes in its facial expressions?" To ensure thorough and meaningful responses, we specifically asked participants to provide the reasoning behind their answers.

3.5 Results

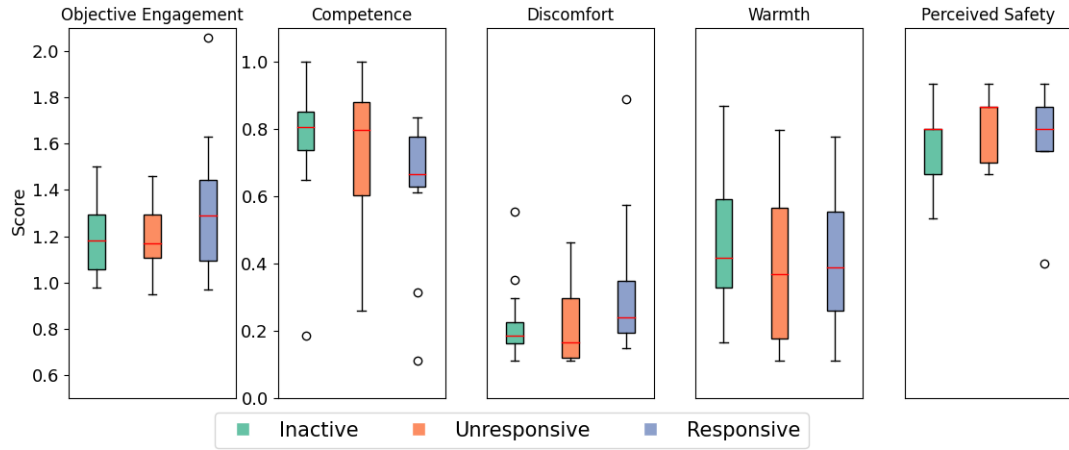
To evaluate participants' perceptions of the robot, we calculated scores for competence, warmth, discomfort, and perceived safety by averaging individual items within each sub-scale. Figure 3.4 illustrates participants' responses for each dimension across conditions using box plots. Each box plot features a line representing the median, with box edges indicating the 25th and 75th percentiles. Whiskers display the range up to 1.5 times the interquartile range, with outliers marked as 'o'.

3.5.1 Variance tests

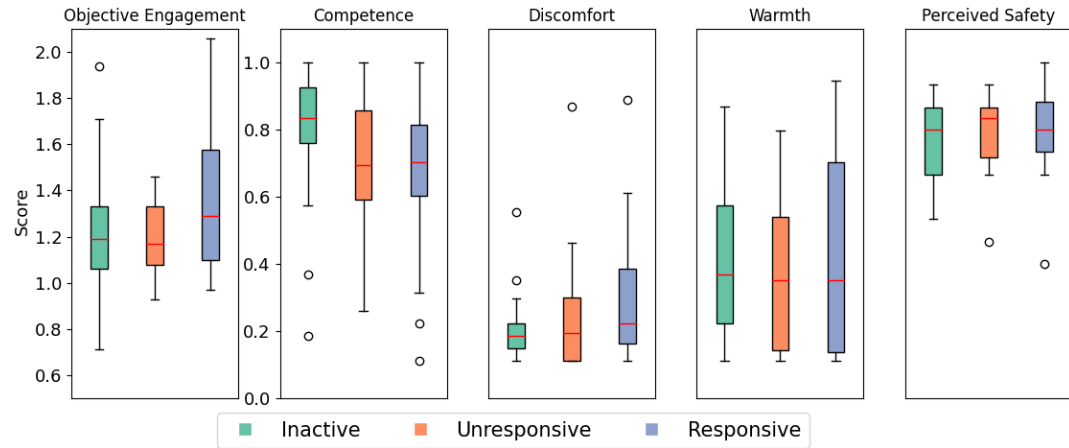
Notably, participants from the older adults group perceived the robot as significantly more competent in the inactive condition compared to the active condition, showing relatively high ratings across conditions, while the younger adults did not show the same trend. Both age groups reported relatively high ratings overall for perceived safety, with no significant differences across all conditions. Ratings of discomfort remained consistently low across all conditions for both age groups, with a trend of higher discomfort in the Responsive condition, though not significant for either of the two age groups. Warmth ratings were comparable between the inactive and unresponsive conditions for both age groups, with relatively low ratings overall. Even if not significant, we can observe that in the older adults group, it is higher in the inactive and unresponsive conditions, while in the younger adults case the unresponsive condition is lower. Additionally, participants consistently expressed high levels of trust across all three conditions, as shown in Figure 3.5a, with a tendency of higher trust in the younger adults group.



(a) Older adults

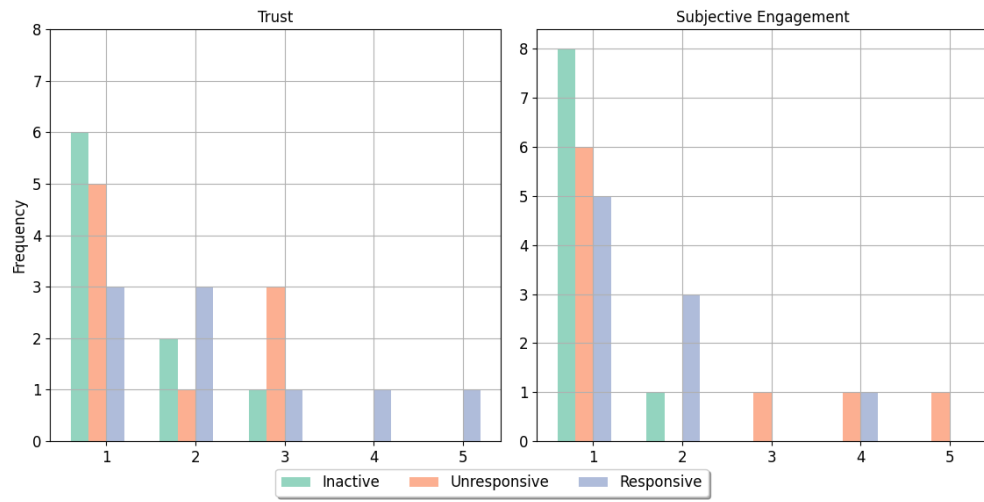


(b) Younger adults

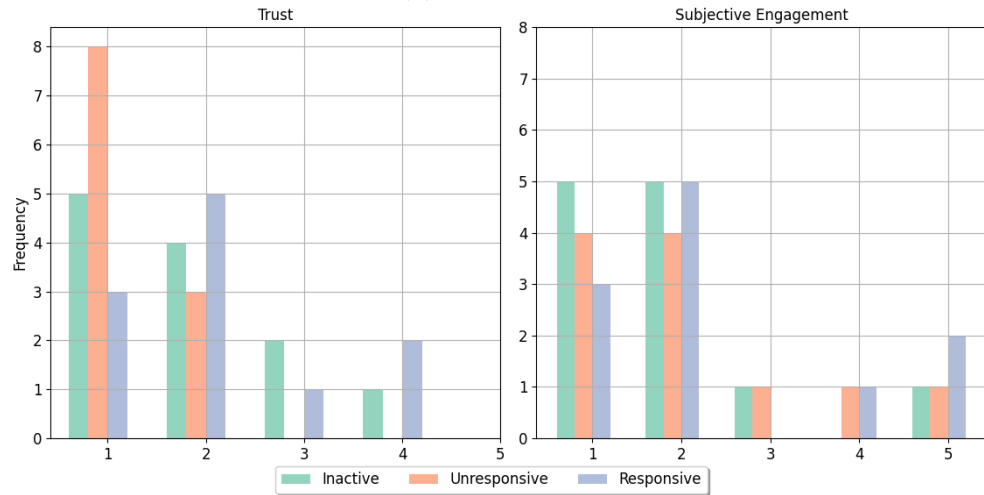


(c) All participants combined

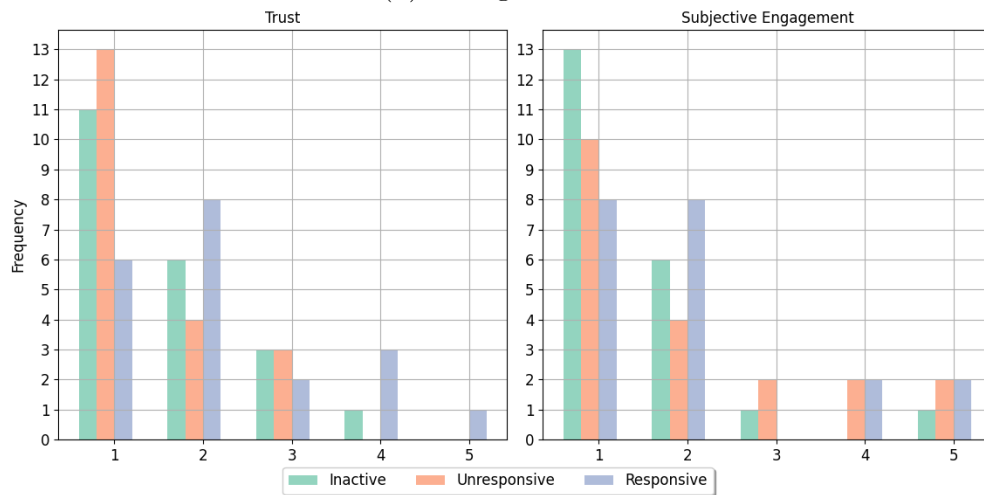
Figure 3.4: Boxplots illustrate older adults 3.4a, younger adults 3.4b, and combined all participants' 3.4c perceived competence, discomfort, warmth, safety, and objective engagement. The ratio E_{obj} is represented as a value potentially exceeding 1. Competence, perceived safety, discomfort, warmth, and safety ratings are normalized to fall within the range of 0 to 1. Significance levels ($** := p - value < 0.01$) are indicated on lines between conditions.



(a) Older adults



(b) Younger adults



(c) All participants combined

Figure 3.5: Bar plots illustrating older adults, younger adults, and combined all participants' perceived trust and engagement for each condition on a 5-point Likert Scale (1 = highest, 5 = lowest).

Concerning engagement, participants consistently reported high levels across all three conditions in both age groups, as shown in Figure 3.5a, with more participants with higher subjective engagement in the younger adults group. During the actual session, participants executed a higher number of pushes, in accordance with the instructions, compared to the expected count calculated from the trial session. The inactive and responsive conditions showed nearly equal ratios for the older adults group, while the unresponsive condition had lower values in comparison to the other two conditions; for the younger adults, the inactive and unresponsive conditions had nearly equal ratios while the responsive condition showed higher values. Objective engagement is presented in the leftmost section of Figure 3.4. Objective and subjective engagement are presented in two separate graphs, as the former is continuous data and the latter is ordinal data.

To analyze significant differences between the three facial expression conditions for each of the age groups, we employed Mann-Whitney U and Chi-Square (χ^2) tests. These non-parametric tests were chosen due to the smaller sample size in each condition ($n < 30$), which precluded a normality check. The selection of tests was based on the type of data being compared – continuous and ordinal. To mitigate false positives, we applied the Bonferroni Correction test. Effect size calculations were performed for significant differences.

For the **older adults** age group, there was a statistically significant difference in competence scores between the inactive condition ($\mu = 8.24, \sigma = 0.55$) and the unresponsive condition ($\mu = 6.07, \sigma = 1.50$), with a $p - value = 0.005$ and $W = 58$. This difference was associated with a substantial effect size of $r = 0.92$. No significant differences were observed for any other dimensions.

For the **younger adults** age group, no significant differences were observed for any dimension.

Furthermore, given the similar trends between the two age groups, we also combined them for a total of $N = 61$ to verify whether there are possible differences across the three conditions as shown in Fig. 3.4c, and 3.5c, but no significant differences were observed for any of the dimensions.

To verify whether there are significant differences in terms of the measured dimensions **between the two age groups**, we also employed Mann-Whitney U and Chi-Square tests, and no significant differences were found for all dimensions between the two age groups.

Answers to RQs sets 1, 2, and 3

From the performed variance tests, our results in response to our first three sets of research questions are as follows:

- **RQ1.1** Facial emotional expressions have a negative impact on older adults' perception of a robot, particularly in terms of competence between inactive and unresponsive conditions. However, these expressions do not influence their level of engagement.
- **RQ1.2** Facial emotional expressions have no significant impact on younger adults in terms of the measured dimensions.
- **RQ2.1 and RQ2.2** Responsive robot facial emotional expressions, in comparison with unresponsive facial emotional expressions, do not have an impact on both older and younger adults' robot perception or engagement levels.
- **RQ3.1** There are no significant differences found between older and younger adults in terms of the impact of facial expressions on engagement and perception, even if competence was significantly different between inactive and unresponsive conditions for the older adults while displaying no significant difference for younger adults.
- **RQ3.2** There are no significant differences in terms of the impact of the responsive facial expression on engagement between older and younger adults.

These outcomes will be discussed in depth in Sec. [3.6](#).

3.5.2 Correlation Analysis

For a deeper understanding, we investigated whether gender exerts an influence on any of the dependent variables. To achieve this, we calculated the Rank- Biserial correlation coefficient (r_{rb}) and conducted Mann-Whitney U and χ^2 tests. Given the dichotomous nature of gender considered in this study, man was assigned the value of 0, and woman was assigned the value of 1 for the numerical analysis. For both the **older adults** and **younger adults** age group, there were no significant correlations ($|r| > 0.6$, $p < 0.05$) between the measured dimensions and gender. However, when combining the two age groups as one single population, a significant correlation was found between gender and competence ($r = 0.58$, $p = 0.0067$) in the inactive condition, indicating that women perceived the robot as more competent than men in this condition.

To enhance our comprehension of the quantitative results, we employed Spearman’s rank correlation coefficient (ρ) to identify correlations within the remaining of the collected data for each of the two age groups, and also for the two groups combined as one single population.

In the inactive condition, as depicted in Figure 3.6a, it is apparent that older individuals with lower stress levels tend to perceive the robot as safer, unlike those with higher stress levels. Conversely, increased engagement with the robot is associated with higher levels of anxiety. These correlations are not reflected in the younger adults group, as we can observe in Figure 3.7a, where a number of significant correlations are uncovered, but none overlap with the ones uncovered for the older adults. Notably, younger individuals displayed that higher stress is associated with higher anxiety, depression, discomfort, and trust, while higher trust is associated with anxiety, depression, and subjective engagement. Subjective engagement is also positively correlated with depression and discomfort. Furthermore, a higher perception of robot competence is associated with higher objective engagement and warmth.

Contrasting the inactive condition, the unresponsive condition reveals intriguing insights for the older adults age group, characterized by a subset of robust correlations ($|\rho| > 0.8$) with an exceedingly low likelihood of arising by chance ($p\text{-values} < 0.005$). Illustrated in Figure 3.6b, participants exhibiting heightened anxiety levels are inclined to perceive the robot as less competent, contrasting those with lower anxiety levels. Additionally, individuals experiencing high stress levels are more inclined to perceive the robot as friendly and warm. For the younger individuals, as shown in Figure 3.7b, once again, the stronger correlations are different with respect to the older adults’ case. Younger individuals with higher anxiety levels also had higher depression, stress, discomfort, subjective engagement, and trust levels. Higher levels of depression were associated with higher levels of stress and subjective engagement, while higher level of stress was also associated with higher discomfort. Individuals who had a higher objective engagement with the robot also perceived it as more warm, while those who perceived it as more competent felt less discomfort.

For the older adults, in contrast to the previously mentioned conditions, the responsive condition exhibited robust correlations ($|\rho| > 0.8$) with dependent variables, demonstrating a high degree of statistical significance ($p\text{-values} < 0.001$). Participants who engage more intensively with the robot tend to perceive it as less warm than those who engage less. Similarly, participants reporting elevated levels of trust perceive the robot as less warm. Also for this condition, the younger adults display different correlations with respect to the older adults, as can be seen in Figure 3.7c. Younger individuals with higher levels of depression were also highly stressed, and engaged more with the robot. In addition, highly

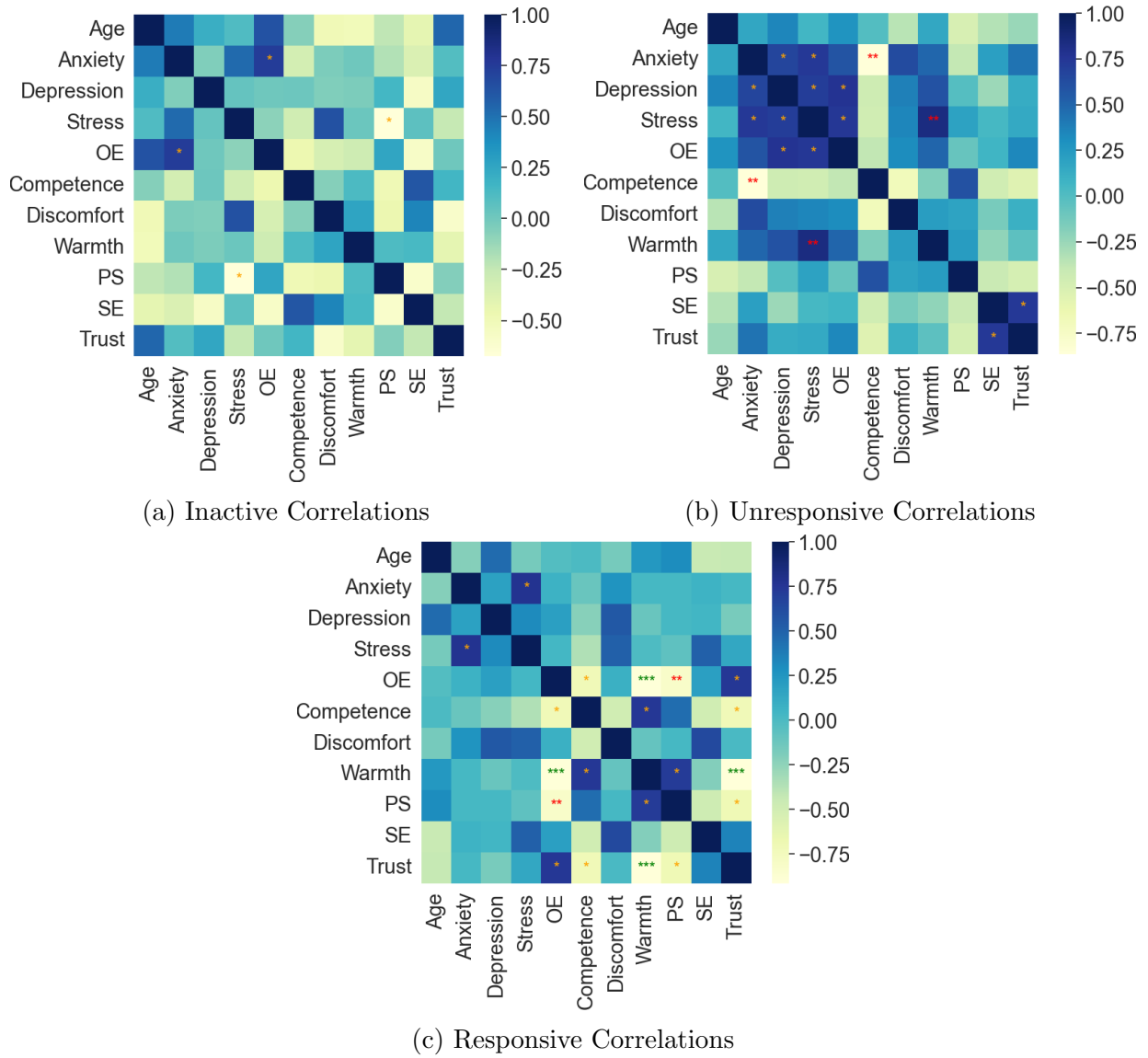


Figure 3.6: Heatmap illustrates Spearman's correlations observed for each condition for the **older adults** age group. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05$, < 0.01 , and < 0.001 with a $|\rho| > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.

stressed individuals also perceived the robot as more competent.

Overall the correlations uncovered were different for the older adults and younger adult

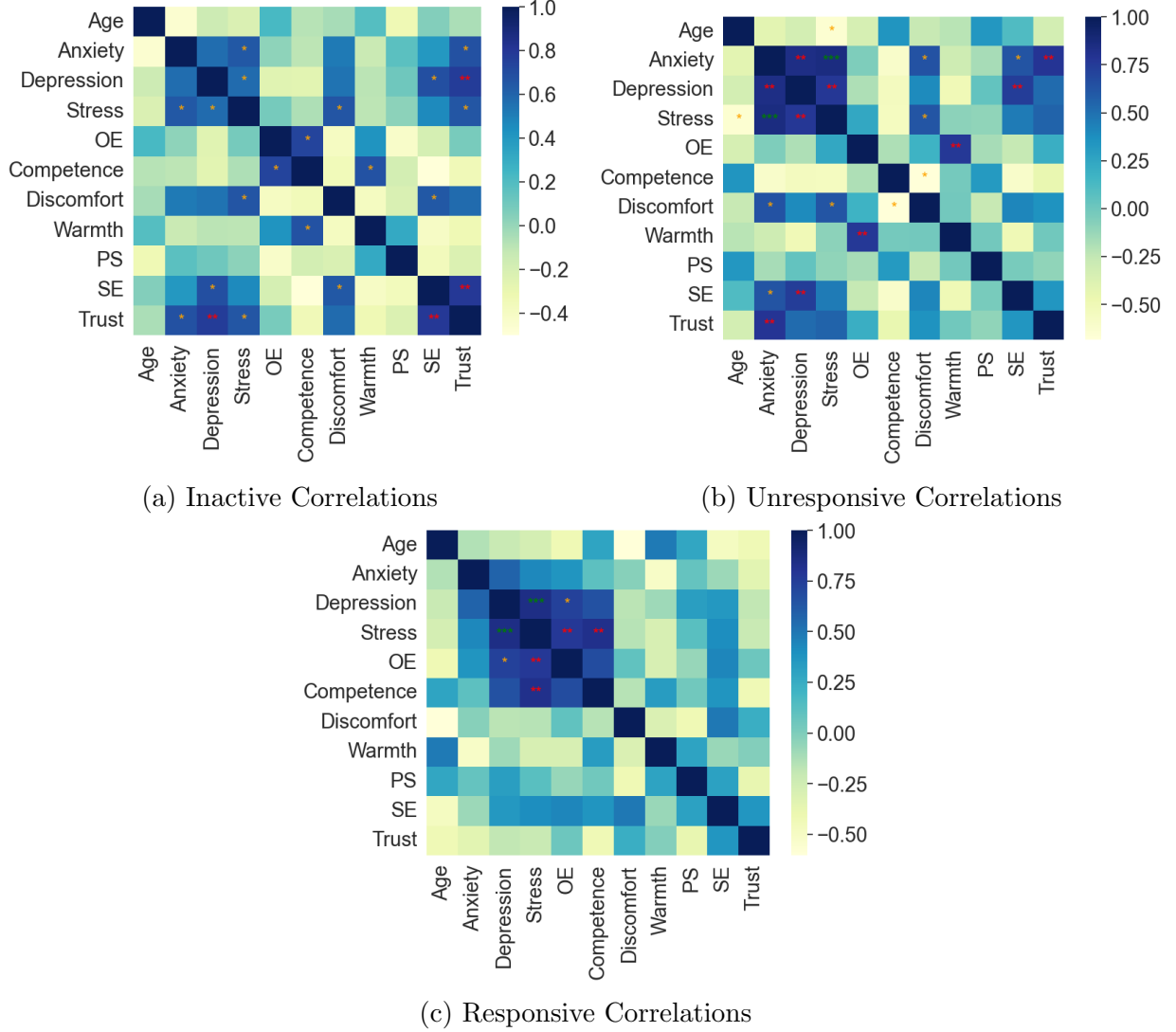


Figure 3.7: Heatmap illustrates Spearman's correlations observed for each condition for the **younger adults** age group. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05, < 0.01$, and < 0.001 with a $|\rho| > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.

populations, with a higher variation in the set of significant correlations across the conditions in the case of older adults, while some of the significant correlations uncovered for the younger adults were consistent across the three conditions. Despite these differences,

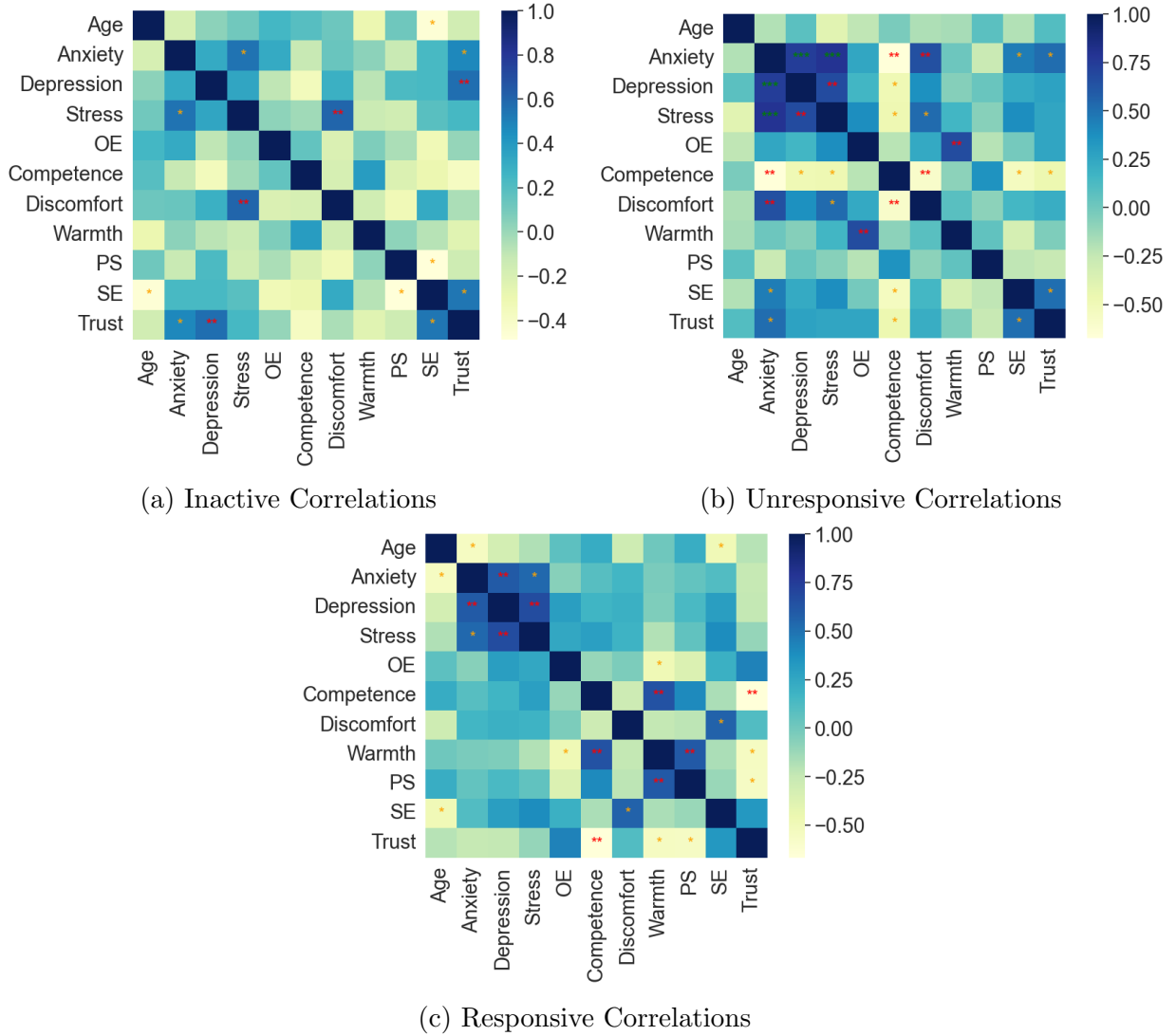


Figure 3.8: Heatmap illustrates Spearman's correlations observed for each condition for the **all the participants**. Correlations marked with '*', '**', and '***' indicate a p -value < 0.05 , < 0.01 , and < 0.001 with a $|\rho| > 0.6$. Abbreviations: OE = Objective Engagement, SE = Subjective Engagement, PS = Perceived Safety.

we also combined the two age groups and conducted the correlations for $N = 61$ to verify whether further insights could be uncovered.

Unsurprisingly, the strong correlations that were identified in each age group mostly

remain, with some being reinforced. Some correlations uncover interesting aspects related to age: for the inactive condition, the older the participants, the lower they evaluated their engagement with the robot; this is also shown in the responsive condition, where in addition, the older the individuals, the lower their level of anxiety. In the unresponsive condition, the strongest correlation are shown for higher anxiety levels associated with higher depression and stress. It is also worth noting that individuals who perceived the robot as more competent also felt the robot as warmer but had lower levels of trust.

It is important to keep in mind that anxiety, stress, and depression levels are assessed prior to the experiment. This means that the measurements reflect participants’ baseline levels in their everyday lives rather than during the course of the experiment itself. Furthermore, no significant differences in anxiety, stress, and depression levels were found between the conditions.

3.6 Discussion

In the following, we first discuss the findings related to the perception of the participants, and then compare directly the older and younger age groups.

3.6.1 Participant perceptions

Our investigation into the impact of facial expressions on the perceived attributes of a robot during a collaborative [pHRI](#) scenario, with an emphasis on older adults, revealed nuanced outcomes. Older adult participants rated the robot as significantly less competent when it displayed facial expressions, contrasting with its inactive state. This trend, while observed among younger adults, did not reach statistical significance. This suggests that facial expressions, in the context of [pHRI](#) like our physical exercise scenario, may not universally enhance robot perception, particularly in terms of competence. This observation can be attributed to the non-social nature of the [pHRI](#) application—physical exercise—employed in our experiment. In contrast, previous literature has shown that incorporating facial expressions into [pHRI](#) applications with a social purpose, such as handshakes, led participants to perceive the robot’s emotions more positively, as evidenced by Tsalamlal et al. (2015) [171]. Furthermore, our results and observations are also supported by participants’ responses to our open-ended questions, in which participant across both age groups referred to the robot as a ”tool” or a ”machine” in both active and inactive conditions. This underscores the importance of future investigations to enhance the robustness of these findings across various [pHRI](#) applications.

Moreover, it is worth considering the distinction between the physical interaction described in Tsalamlal et al.’s study [171], where a humanoid was employed, and our utilization of a collaborative robot like Sawyer, equipped with a screen. This contrast prompts a pertinent inquiry regarding how the type of robot employed in pHRI might influence robot perception. It is conceivable that people’s preferences for engaging in physical interaction and communication could differ when interacting with a humanoid displaying facial expressions compared to a collaborative robotic arm with a screen conveying facial expressions. However, we believe that the context, whether social or collaborative, significantly influences user communication preferences.

Additionally, a notable recommendation arises to enhance the robot’s facial features by incorporating elements such as cheeks or eyelids, aligned with the findings from Chen and Jia’s study [30], thereby imparting a sense of maturity and user preference to the robot’s appearance. This recommendation finds additional support through a considerable number of responses obtained from participants in our study’s open-ended questions, wherein many participants suggested altering the robot’s facial features. One participant even specifically mentioned the eyes. Implementing this adjustment holds the potential to cultivate a perception of Sawyer as more mature and visually appealing.

The study revealed a notable discrepancy in competence levels between the inactive and unresponsive states among older adults. However, contrary to expectations, this disparity was not mirrored when comparing the responsive condition. In this scenario, competence levels in older adults were similar to those observed in the unresponsive state, but no statistically significant difference was found compared to the inactive condition. This pattern was also observed among younger adults. Given these trends, it is possible that significant differences may be uncovered with a larger sample size for all age groups.

We speculate that the absence of a significant difference between the responsive and unresponsive conditions in both age groups might be attributed to the abstract nature of the chosen set of facial expressions and the specific type of application employed in this experiment. This conjecture is supported by the fact that 50% of all participants in the responsive condition reported not perceiving any changes in the robot’s facial expressions, as their attention was directed towards activating the headlight blinks, this was especially true in the case of older adults, where 4 out of 9 participants did not notice the change, while in the case of younger adults, only 3 out of 11 did not notice. This perspective gains further weight from one participant’s response to the open-ended questions, where they suggested altering the facial expressions once the exercise was completed. This observation underscores the possibility that the nature of the application and its objective could have influenced participants’ focus on facial expressions. It further implies that pHRI applications primarily centered around collaborative tasks, as opposed to sociability, such as

engaging in physical exercises instead of handshaking, should consider incorporating facial features that are less abstract and more conspicuous.

The absence of a significant difference in both age groups’ participants’ levels of engagement across all conditions suggests that facial expressions may not exert any influence on engagement during specific types of [pHRI](#) applications. However, an argument can be made that the robot’s joint impedance control inhibited certain participants from executing a high number of pushes, as they waited for the robot’s arm to return to its initial position after each push. In contrast, other participants proceeded without such hesitations. This divergence in exercise approaches highlights a potential limitation: the count of pushes might not have been the most suitable metric for assessing objective engagement in this context. Therefore, we encourage researchers to explore better-suited engagement measures.

Similarly, the lack of statistically significant differences in the subjective engagement measure bolsters the notion that facial expressions might not significantly impact participants’ levels of engagement. This interpretation gains further support from the notable proportion of participants who recommended additional features like rhythmic music, encouraging phrases, visual timers, and verbal motivation. Collectively, these inputs suggest that facial expressions might not be the most suitable non-verbal social cue for enhancing engagement in collaborative [pHRI](#) applications.

The correlation analysis was performed to further support our research findings. The correlations between participants’ baseline emotional states—specifically anxiety and stress levels—and their perceptions and engagement with the robot offer significant insights. This underlines that emotions are not peripheral to [pHRI](#); they are integral. Notably, the analysis reveals that individuals with lower stress levels tend to perceive the robot as safer, suggesting that a relaxed state may predispose participants to view the robot more favorably in terms of safety. Conversely, those with higher levels of anxiety perceive the robot as less competent, indicating that pre-existing anxious dispositions could negatively impact the perception of the robot’s ability to perform tasks effectively.

The varying correlations seen in different settings highlight how subtle differences in the experiment can significantly affect how participants perceive and experience the interaction. For instance, in conditions where the robot exhibited facial expressions, some participants did not perceive changes in these expressions, potentially because their attention was focused on other aspects of the interaction, such as completing the physical exercise task. This observation suggests that the effectiveness of facial expressions in enhancing robot perception and engagement may be contingent upon the context of the interaction and the participants’ focus during the task.

Furthermore, the analysis highlights the role of engagement in shaping perceptions of

the robot. Participants who engaged more intensively with the robot tended to perceive it as less warm in the responsive condition, suggesting that high levels of task engagement might overshadow the social cues provided by the robot’s facial expressions. This finding points to the importance of balancing task-oriented and social interaction elements in designing effective [pHRI](#) experiences.

The correlations between participants’ gender and perceptions of the robot, although not a primary focus of our analysis, suggest that demographic factors could also influence interaction outcomes. Future research could explore the impact of gender and other demographic variables on [pHRI](#) experiences in more depth.

The correlation analysis highlights the importance of considering participants’ emotional states and demographic characteristics in designing and evaluating robot interactions. These findings suggest avenues for future research, including investigating the optimal balance between task-oriented and social interaction elements in [pHRI](#) and exploring the impact of demographic factors on interaction outcomes.

3.6.2 Comparative analysis of older and younger adult age groups

Despite the lack of significant differences between the older and younger adult age groups when directly compared via variance tests, our results unveiled interesting observations that highlight how age can differently affect how people view and interact with robots during physical human-robot interactions. Interestingly, the impact of robots’ facial emotional expressions on participants’ perception of the robot varied significantly between the two age groups, particularly regarding the robot’s competence. Older adults perceived the robot as less competent when it displayed facial expressions compared to when it did not, a trend that was observed to a lesser extent among younger adults but did not reach statistical significance. This divergence in perception suggests that older adults might prioritize different attributes in robots than younger adults, possibly valuing functionality or the ability to complete tasks over social or emotional capabilities. As a matter of fact, some of the older adult participants have commented that the older generation may see robots just as machines and not as social agents, whereas younger adults who have been brought up surrounded by digital technologies that may act as artificial agents (e.g. smartphones, smart speakers) might have a different view on robots and their social role. However, this may also depend on the robot’s physical appearance and designated tasks [106]. It is also worth noting that people’s attitudes towards digital technology have also significantly shifted over time. In the 90s, computers were perceived as social actors (Computers are Social Actors - CASA theory), however recent studies have shown that this theory does

not hold anymore [72], so we may hypothesize that the same shift could happen for robots, though the complexity of this assessment would be much higher compared to computers.

Engagement levels across all conditions did not significantly differ between older and younger adults, indicating that facial expressions might not be a pivotal factor in determining engagement within the specific types of pHRI applications explored in our study, such as physical exercises. This finding is particularly intriguing as it suggests that other factors, perhaps related to the design of the interaction or the physical activity itself, play a more critical role in fostering engagement across age groups.

Moreover, our study found no significant differences in perceptions of warmth, safety, or discomfort associated with the robot across the different facial expression conditions for either age group. This lack of significant difference between responsive and unresponsive facial expressions further indicates that the dynamic nature of facial expressions may not significantly alter participants' perceptions within the confines of our physical exercise scenario, regardless of age.

The correlation analysis highlighted how individual differences in anxiety and stress levels might influence perceptions of the robot's attributes, such as safety and competence, and engagement levels with the robot, across both age groups. This suggests that participants' baseline emotional states could play a role in shaping their interactions and experiences with robots in pHRI scenarios. However, the specific correlations observed differed between older and younger adults, notably younger adults displayed significant correlations between the measured dimensions with anxiety, stress, and depression levels. We could speculate that this could be due to the difference in state of mind and societal roles between older and younger adults. All of the recruited older adults were retired and living an independent life, while the recruited younger adults were mostly university students. This difference indicates the complexity of designing robots that can effectively engage and be perceived positively by diverse user groups.

3.7 Limitations

We acknowledge that there was a set of limitations in this study that may have undermined some of the findings. First of all, a larger sample size could have potentially revealed more statistically significant differences between dimensions, reducing the risk of Type II errors. However, due to the unique and challenging nature of recruiting older adults, obtaining a larger sample size was unfeasible. Our sample size is greater than previous studies involving robots and older adults, such as Nowak et al. [124], who could only recruit 7 participants, and Giorgi et al. [60], who recruited 17 participants to investigate perceived

trust in a social robot across a wider age range (40 to 87 years). Moreover, the study’s design, conducted as a between-subject experiment to mitigate issues highlighted in prior literature [49], unintentionally accentuated differences between conditions, reinforcing the need for a greater participant count.

A significant aspect of this study’s methodology is its reliance on a quantitative measure of engagement, specifically the number of physical interactions (pushes) against the robot. While this approach provides a tangible metric for engagement, it potentially oversimplifies the multifaceted nature of human-robot interactions. Engagement with robots, especially in contexts blending physical activity with social interaction, is likely influenced by a range of factors beyond mere physical contact, including emotional connection, perceived empathy, and social presence. Therefore, a more holistic approach to measuring engagement, incorporating both qualitative assessments and possibly physiological indicators, might reveal deeper insights into the dynamics of human-robot relationships.

The study’s use of abstract facial expressions displayed on a screen to convey robotic emotions represents another pivotal limitation. Given the intricate ways humans interpret and react to emotional cues, the abstract representations used here may not fully align with participants’ expectations or the subtleties of human emotion. This discrepancy highlights the importance of exploring more realistic or nuanced facial expressions, potentially leveraging advances in robotics and artificial intelligence to create robots capable of more authentically mirroring human emotional states.

Furthermore, the physical design of the robot—a collaborative robotic arm equipped with a screen—along with the specific context of physical exercise, likely influenced participants’ perceptions and interactions. This setup, while innovative, might not lead to the same level of social connection or empathy as interactions with more humanoid robots or robots designed with a clear emphasis on social engagement. The study’s insights, therefore, may have limited applicability to scenarios where the robot’s physical form and the nature of the task differ significantly, underscoring the need for research across a broader spectrum of robot designs and interaction contexts.

Subjective assessments of the robot’s competence, warmth, and safety, obtained from participant responses to questionnaires, introduce another layer of complexity. Such perceptions are deeply personal and can vary widely based on individual age, background, experiences, and expectations, making it challenging to find generalizable trends or insights.

Last but not least, much of the foundational research in HRI predates the COVID-19 pandemic, with recent studies (after 2021) often building upon these earlier works. However, the pandemic has dramatically changed how people view and use technology,

including robots. There's a noticeable increase in openness towards using digital and robotic solutions [59, 176]. This shift could have influenced how participants in recent studies interact with and think about robots, suggesting we're moving into a new phase of technology acceptance. Comparing these latest findings to older studies might not be straightforward, as the pandemic has changed the playing field, highlighting the need to consider these new attitudes towards technology when evaluating HRI research.

3.8 Conclusion

Our research aimed to uncover the impact of facial expressions on a robot's perceived attributes and an individual's level of engagement within a collaborative [pHRI](#) application, by focusing on two age groups independently and comparatively. We found that older adults perceived robots displaying facial expressions as less competent, a perception not significantly shared by younger adults, indicating that facial expressions do not uniformly enhance robot perception. Additionally, facial expressions did not significantly affect engagement levels for any age group, suggesting that other aspects of interaction design or the activity itself are more influential in promoting engagement. Statistical analysis, including variance tests and correlation analysis, helped us examine how participants' baseline emotions affect their interaction with robots, highlighting the importance of individual differences in [HRI](#) experiences. The insights gained from this process can guide future research in this area, ensuring that we can effectively capture the complexities of human-robot interactions.

The limitations of this study outline possible future research in [pHRI](#) and in our specific scenario. Expanding sample sizes, employing improved measures of engagement, exploring the effects of more realistic facial expressions, considering diverse robot designs, and conducting studies across varying contexts are all directions to improve our understanding of [pHRI](#).

Chapter 4

Assessment of Robot Perception Subjective Quantification Approach

4.1 Introduction

As discussed in earlier chapters, assessment of subjective tools is crucial for identifying appropriate objective quantification methods through the determination of robust and statistically significant correlations. Therefore, this chapter is dedicated to analyzing the data obtained in Chapter 3 to assess the reliability of the selected tool. This analysis shall help us understand whether the chosen approach is suitable for discerning appropriate objective quantification methods.

The Godspeed questionnaire [15] has been regarded as the gold standard for subjectively quantifying robot perception in HRI. Comprising five subscales rated on a 5-point semantic differential scale—anthropomorphism, animacy, likeability, perceived intelligence of the robot, and perceived safety—it has enjoyed widespread use. However, despite its popularity, several studies have highlighted pitfalls in the Godspeed questionnaire in HRI [29], and in direct pHRI scenarios [76] in particular. To address these limitations, Carpinella et al. [29] developed RoSAS, drawing inspiration from both the Godspeed Scale and psychological research on social perception. RoSAS comprises three subscales rated on a 9-point Likert scale—competence, warmth, and discomfort. Their statistical analysis revealed that competence and warmth measured by RoSAS align closely with the likeability and perceived intelligence dimensions of the Godspeed Scale. However, they also noted that RoSAS does not comprehensively cover all dimensions of the Godspeed Scale. Therefore, they recommended supplementing RoSAS with relevant Godspeed constructs based on specific study

needs. Therefore, in our studies, we have adopted this approach by assessing robot perception using [RoSAS](#) and complementing it with perceived safety from the Godspeed Scale. Perceived safety is particularly important as it encompasses fundamental dimensions in both ergonomics and hedonomics [1].

Furthermore, [RoSAS](#) is primarily developed online with robot faces, prompting authors to stress the importance of evaluating the reliability of this questionnaire in various [HRI](#) scenarios with physical robots. However, insufficient effort has been directed towards validating and evaluating [RoSAS](#) in [pHRI](#). One notable attempt in this regard is the evaluation of the internal consistency of [RoSAS](#) in a handover scenario [133]. While [RoSAS](#) has shown high reliability for indirect [pHRI](#) scenarios such as handovers, there is a lack of literature regarding its suitability for direct [pHRI](#) scenarios.

4.2 Method

Since the variance tests conducted showed no significant differences between the two age groups regarding the dimensions measured with [RoSAS](#) and perceived safety, the two groups were merged into a single population ($N = 61$). This consolidation allowed for the validation of the [RoSAS](#) and perceived safety questionnaires independently for the three conditions. In [HRI](#), questionnaires are frequently assessed using Cronbach’s alpha [15, 29, 133], which serves as a measure of internal consistency (reliability) [169]. It is essential to identify the unidimensionality of scales before calculating Cronbach’s alpha, as it is presumed. Unidimensionality can be assessed using factor analysis. To determine the suitability of factor analysis, the [KMO](#) test [40] is typically conducted. Only after confirming the suitability of measurable dimensions for factor analysis and determining unidimensionality are outliers removed from the dataset at the score level to prevent any false inflation of Cronbach’s alpha due to asymmetry [98]. Furthermore, Cronbach’s alpha is calculated after omitting some items to assess whether any items may need to be excluded from the questionnaire.

As detailed in Chapter 3, we aimed to recruit a substantial number of participants from diverse age groups to ensure a robust assessment of the subjective quantification measure.

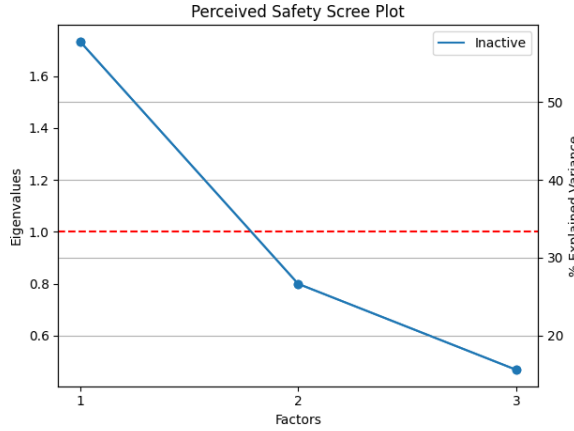
4.3 Results

Most dimensions demonstrated suitability for factor analysis, unidimensionality, and consistent reliability across the three facial emotional conditions. However, perceived safety

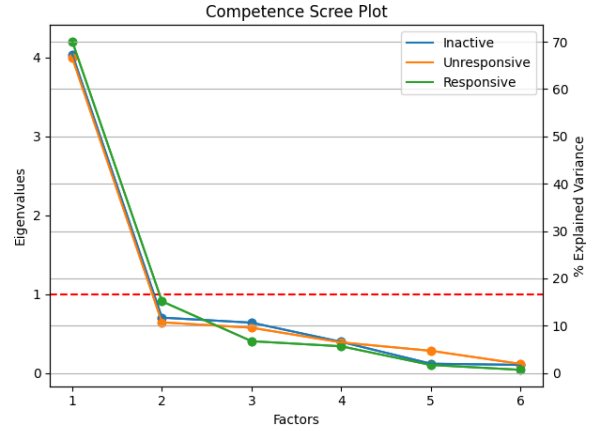
across the active conditions did not exhibit suitability for factor analysis, with a **KMO** value below 0.5, whereas the inactive condition showed a **KMO** value of approximately 0.6. In contrast, the remaining dimensions displayed a **KMO** value of 0.6 or higher, indicating perfect suitability for factor analysis. Considering the acceptable **KMO** value for the inactive condition in perceived safety, it was included in the factor analysis with the other dimensions.

All dimensions demonstrated unidimensionality based on the calculated eigenvalues, as illustrated in Figure 4.1. A higher eigenvalue indicates that an item explains a greater amount of variation. For an attribute to be considered unidimensional, a single item would explain a substantial part of the attribute's variance, typically indicated by an eigenvalue exceeding 1. This means that for all four attributes, one single factor explains most of the variances, indicating that the attributes are unidimensional.

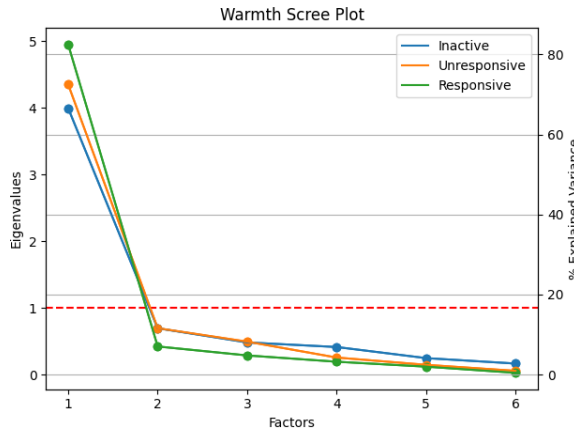
The percentage of explained variance for the inactive condition of perceived safety was 57.77%, while competence exhibited a percentage of explained variance ranging between 66.5% and 70% across all conditions. Similarly, warmth showed a percentage of explained variance between 66.5% and 82.5% across all conditions, and discomfort ranged between 59% and 74.5% across all conditions.



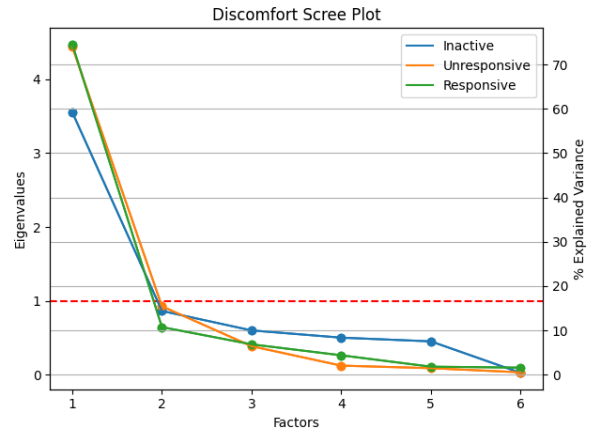
(a) Perceived Safety



(b) Competence



(c) Warmth



(d) Discomfort

Figure 4.1: Scree plots of the factor analysis performed on perceived safety (Godspeed) for the inactive condition only based on the **KMO** values, and the three dimensions of **RoSAS** (competence, warmth, and discomfort), for all three conditions.

After removing outliers, Cronbach's alpha for perceived safety in the inactive condition initially showed a very low value of 0.13. However, upon removing the item 'Quiescent:Surprised', the alpha increased to 0.66. Competence exhibited very comparable alphas above 0.7 across the three conditions, while warmth showed even higher alphas around 0.9 across all three conditions. In contrast, discomfort showed considerable diversity in alphas across the three conditions, with the inactive condition displaying a very low alpha of 0.2, the unresponsive condition of 0.77, and the responsive condition of 0.87. Further-

more, omitting the item 'Strange' from discomfort resulted in an increase in the inactive condition, but had minimal impact on the active conditions.

4.4 Discussion

All dimensions demonstrating unidimensionality also displayed a substantial percentage of explained variance, suggesting that the selected items adequately represent the variation within each dimension. Additionally, most dimensions exhibited acceptable alpha values.

The high correlation observed between the items 'Anxious:Relaxed' and 'Agitated:Calm' in the perceived safety results suggests that substituting the 'Quiescent:Surprised' item with a more suitable alternative might increase the alpha coefficient. We anticipate that the unfamiliarity with the term 'Quiescent' may have contributed to the low alpha score, as evidenced by some participants (native speakers) inquiring about its meaning. This adjustment could potentially improve internal consistency, as the number of items in a subscale affects alpha values. Regarding competence and warmth, the items demonstrate robust reliability and can effectively quantify these dimensions, capturing the variation within each dimension adequately.

In terms of discomfort, while one might claim that it demonstrates acceptable internal consistency, the notably low alpha coefficient in the inactive condition casts uncertainty on this conclusion. It is possible that the discomfort subscale may not be suitable for scenarios involving human-robot interactions without faces, as [RoSAS](#) is developed based on evaluating robot faces. However, [\[133\]](#) have demonstrated good Cronbach's alpha for the discomfort dimension in a handover scenario with a serial manipulator. Therefore, further efforts are needed to thoroughly understand the internal consistency of this measure.

While Cronbach's alpha is commonly used to assess reliability and internal consistency in subjective quantification approaches, it is important to note that high alphas do not always indicate a high level of internal consistency, as they are influenced by factors such as the number of items in each subscale [\[169\]](#). Short tests tend to yield lower alphas. However, in our case, with each item represented by at most 6 items, we argue that our alphas do not suffer from such a limitation.

Moreover, it is crucial to recognize that alpha is specific to the sample of participants and should be measured each time the test is administered, rather than relying solely on alphas from other scenarios. Additionally, while evaluating questionnaire reliability is necessary, it does not validate the questionnaire. Construct and criterion validity can only be confirmed through numerous studies involving a substantial number of participants.

Therefore, we encourage the [HRI](#) research community to measure and report the internal consistency of their subjective tools, as this is essential for confirming validity.

In summary, the [RoSAS](#) questionnaire emerges as a reliable subjective measure for evaluating robot perception within direct [pHRI](#) scenarios. However, there remains uncertainty regarding the suitability of perceived safety as a representation of robot perception. The low [KMO](#) values observed across two out of the three conditions in perceived safety suggest that alternative approaches to factor analysis for determining unidimensionality might yield more insightful results. Therefore, we cannot definitively assert the unsuitability of the perceived safety dimension within the robot perception measure.

It should be noted that while not all dimensions exhibit high internal consistency, they can still serve our purposes. Developing a questionnaire is a time-consuming process that often relies on online studies due to practical constraints on participant recruitment for real-world experiments [29]. Attempting to recruit a large number of participants for such experiments would be impractical and could divert us from our research goals. Therefore, we believe that accepting a slight uncertainty in internal consistency in some dimensions is a reasonable trade-off for achieving our overarching objectives.

4.5 Aligning with Thesis Objective

While internal consistency is essential in all experiments, our focus in Chapter 3 was to gauge the subjective tool’s effectiveness in a simple task before applying it in more complex scenarios. This approach, akin to a sanity check, allowed us to recruit a diverse participant pool across different age groups and ensured a reliable assessment of the subjective quantification measure. Assessing the tool in a complex setting poses the risk of undermining any correlations found between subjective and objective measures. Chapter 5 further discusses the challenges encountered in participant recruitment for more demanding tasks.

Based on the findings in this chapter, we choose to utilize [RoSAS](#) in conjunction with perceived safety for subjective quantification of robot perception in direct [pHRI](#) scenarios. However, it is important to note that perceived safety and discomfort results should be interpreted with caution.

Chapter 5

Quantifying Human Mental State in Interactive pHRI: Maintaining Balancing

5.1 Introduction

pHRI plays a crucial role in supporting independent living for older adults, especially in scenarios requiring complex physical assistance. In these situations, the robot may need to apply unforeseen forces while collaborating on tasks like showering, and transferring. While users anticipate the robot’s involvement in these tasks, the specific manner and frequency of force application may not be expected at every step. Additionally, there are instances where the robot must intervene physically without prior notification to prevent a person from losing balance. Therefore, understanding people’s perceptions of these unanticipated actions by caregiving robots is crucial.

While *passive* pHRI refers to compliance of the robot to the user motion while physically coupled, we define *active* pHRI as unanticipated physical actions initiated by the robot while physically coupled [76].

As previously mentioned, robots need to understand when a user is uncomfortable, stressed, or anxious when being handled in a certain way. Real-time prediction of an older adult’s mental state is essential for effective pHRI scenarios. This prediction can be achieved by combining real-time data, such as skin temperature, heart rate and body posture, with demographic and personality information gathered beforehand. According to the literature, this approach can effectively contribute to predicting a user’s mental

state during pHRI scenarios [75]. However, further investigation and data gathering is necessary to develop a mature human mental state model that can accurately predict the user’s mental state in real-time during pHRI scenarios.

As aforementioned in Chapter 2, Section 2.9, we identified cognitive ergonomics and belief as the two human factor types associated with cognitive aspects that best represent human mental state. Cognitive ergonomics involves designing systems that align with users’ perceptual and psychological abilities, while belief refers to cognitive responses that can trigger emotions (i.e., affective responses). Some of the measurable dimensions that belong to belief and cognitive ergonomics include robot perception; and mental workload and stress, respectively. To accurately measure each dimension, we adopted a combination of objective and subjective approaches whenever feasible. Objective measures encompassed PPG, skin temperature, and task performance, while subjective assessments involved utilization of standardized questionnaires from domains beyond robotics due to the absence of an accurate subjective questionnaire tailored to our needs. Additionally, usability of the system were assessed for better understanding of the topic. More details are indicated in Section 5.4

Our study aims to identify the effect on the user’s mental state due to different robots’ behaviour (active vs passive pHRI) and contribute to the development of a human mental state model by correlating potential real-time measurable data (objective quantifiable data) and pre-gathered data to human mental states. To this end, we adopt a stabilizing task as a common application in caregiving scenarios, where users can utilize the robot to maintain balance. Our findings are expected to provide valuable insights into the development of robots with advanced decision-making abilities and the capability to adapt to users’ mental states. This could pave the way for more sophisticated and user-friendly robots, opening up new possibilities for human-robot interaction and collaboration.

This chapter is structured as follows: Section 5.2 provides an overview of existing research on objective quantification of human factors in direct pHRI scenarios; in Section 5.3, the research questions guiding the study are presented; Section 5.4 details the experiment design, participants, and measures taken for evaluation purposes; Sec. 5.5 illustrates the obtained results; Section 5.6 discusses the obtained results with respect to the outlined research questions, and outlines the limitations of the presented study, providing insights for possible future improvements and developments; Section 5.7 concludes with a brief summary and take-aways.

5.2 Related Work

Remarkable efforts have been made to study a wide range of human factors in various [pHRI](#) applications, including physical exercise [52], hugging [37], and dancing [132]. However, these studies have primarily adopted situations where participants expected the robot to apply force against them. Few studies have been conducted to investigate active [pHRI](#), where participants do not expect the robot to apply force against them. Moreover, there have been only a few studies that have examined the correlations between human mental states and either pre-gathered or real-time measurable data.

Haring et al. [68] examined the impact of social touch on the human perception of a female-looking android robot. They explored the relationship between participants' perception of the robot and their personality traits, as well as physical measurable data such as the total touch duration and the time taken before approaching the robot's hand. The study found that participants with higher levels of neuroticism perceived the robot as more human-like and animated, resulting in hesitation to touch it. In contrast, extroverted participants perceived the robot as less human-like and less animated, leading to quicker approaches. Furthermore, the authors observed that participants with higher levels of extroversion tended to use more forceful actions, such as squeezing the robot's hands, while those with higher levels of neuroticism tended to interact more gently with the robot, using actions such as stroking.

According to a study by Bainbridge et al. [14], individuals who hold more positive opinions towards shaking hands with a humanoid robot tend to exhibit higher hand temperatures, higher tactile measurements, higher temperature growth, and farther face distances. The study found that temperature has the strongest and most stable correlations with participants' perceptions, followed by temperature growth and face distance, then tactile measurements, and finally force measurements. In contrast to Bainbridge et al. social scenario, Muller et al. [117] conducted a [pHRI](#) study in which participants manipulated an industrial robotic arm to move a set of weights from one place to another. Their interesting findings indicate that comfort is positively correlated with following the participants' intentions, feeling of mass, precision, and self-efficacy. Furthermore, they discovered that individuals with low neuroticism are generally faster at completing the task, while those who are prone to stress are slower, which is consistent with [68] findings with an android robot. This raises the question of to what extent correlations between personality traits and physical data, such as the time duration in touching or manipulating a robot, can be generalized across completely different types of robots, such as an android robot and a relatively big-sized industrial manipulator.

As indicated in chapter 2, a significant amount of research has been conducted to de-

termine whether human mental states can be predicted based on real-time measurable data. Some studies utilized machine learning models [44], while others exploited mathematical models [175]. However, most of these studies involve a high degree of complexity in predicting human mental states.

Given the relative youth of pHRI research, no wonder that only a limited number of studies have investigated active pHRI. Hu et al. [76] conducted an exploratory study in a gaming context to identify measurable interaction factors that could be useful for developing active pHRI controllers. This study was the first attempt to investigate the impact of active pHRI, and its main contribution is providing insights and highlighting potential pitfalls that researchers should be aware of when conducting active pHRI studies.

The study stresses the importance of ensuring that participants are motivated and interested enough to take part in the collaborative task with the robot. This is essential to identify significant differences between passive and active physical interactions. Additionally, researchers should ensure that any unanticipated forces applied by the robot are noticeable to participants, without being mistaken for a malfunction.

While the study did not reject the null hypotheses, it had some interesting findings. The researchers identified potential factors that have a significant impact on how participants perceive active pHRI. For example, they found that extroverted individuals tended to approach the robot from a closer distance. Furthermore, individuals who perceived the robot as more anthropomorphic and animated showed higher levels of stress and anxiety during active pHRI.

Hu et al. [75] conducted an exploratory study to investigate how participants perceive and emotionally react to a robot in a gaming context, where the robot is controlled like a joystick to catch virtual coins. Unbeknownst to the participants, the robot was programmed to subtly guide them toward high-value coins, resulting in higher scores and greater monetary rewards. The study found that users' perceptions, feelings, and physical data were significantly impacted by the robot's unanticipated physical actions.

During active pHRI sessions, participants showed increased interaction forces and distance between their feet compared to passive sessions. Moreover, interacting with the robot in active pHRI mode led to a decrease in pleasure and an increase in arousal and dominance. The authors were able to correlate participants' feelings and perceptions of the robot to real-time measurable data and pre-gathered data, creating clusters based on personality traits, body posture, and robot perception.

Their results indicate that interaction forces were strongly linked to age, personality, and robot perception, while personality was strongly correlated with body posture and robot perception. Participants who had a positive attitude towards the robot tended to

interact with it in a more relaxed manner, with closer proximity—contrary to the findings of [14]—and applying lower forces. Moreover, they demonstrated lower levels of anxiety (as indicated by higher [Galvanic Skin Response \(GSR\)](#) compared to baseline) and reduced mental workload (evidenced by smaller pupil dilation). Participants who understood that the robot was trying to be helpful, used less force, experienced less mental load, felt more dominant, and kept both hands closer to the robot.

The authors also found that individuals with higher levels of extroversion preferred closer proximity to the robot, confirming their previous research [76]. However, they were unable to establish a correlation between gender and real-time measurable data or human mental state.

Building on previous studies, our current study aims to investigate the impact of active [pHRI](#) on people’s mental state during a physical assistance task. To the best of our knowledge, no similar study has been conducted in this area. Additionally, we aim to contribute to the development of a human-state model that can be used as a predictive tool for future active [pHRI](#) controllers.

5.3 Research Questions

It was evident from the preceding section that there have been only a few studies that have addressed active [pHRI](#) scenarios. Moreover, studies examining active [pHRI](#) have often used a large number of sensors (dependent variables) relative to the number of participants, mostly in the context of gaming [75]. Thus, our objective is to enhance the current understanding of the effects of active [pHRI](#) by conducting a study that represents a near-realistic scenario, using a minimal set of sensors and a feasible number of variables for analysis, and an adequate number of participants. Furthermore, we aim to examine an assistive [pHRI](#) application, addressing the literature gap in unexplored collaborative applications of [pHRI](#) [1].

Additionally, only a limited number of studies have explored the correlations between potential real-time measurable data, demographic information, or personality information and the human mental state in [pHRI](#) scenarios. Therefore, our study aims to advance the development of a human-state model by determining the correlations between the participants’ mental states toward the robot and both real-time physical data and pre-existing demographic and personality information. To achieve this, we have formulated the following research questions:

RQ1: Will an active robot behaviour impact the mental state of participants compared to passive robot behaviour in an assistive [pHRI](#) scenario?

RQ2: What correlations exist between participants’ demographics, personality traits, physical data, physiological data, and their mental state during a physical assistive [pHRI](#) scenarios?

For **RQ1**, our goal is to understand how a robot’s active behavior affects human mental state during collaboration, leading to a deeper understanding of this impact. Additionally, addressing **RQ2** will contribute to developing a human-state model for predicting mental states during [pHRI](#). This model will support the creation of adaptive controllers that factor in human mental states when making decisions during [pHRI](#).

5.4 Methodology

As a common application in caregiving scenarios, we have chosen a balancing task as the appropriate methodology to address our research question. The experiment took place in the Active & Interactive Robotics Laboratory at the University of Waterloo, Ontario, Canada.

Participants were instructed to stand on a perturbed balance board, simulating a sensation of imbalance without risking actual falls. Figure 5.1 show the front and side view of the experiment setting. They were then tasked with maintaining their balance, with the option to hold onto the robot for assistance. The stabilizing task involved two conditions:

1. **Passive behaviour**¹ : The robot complies with the participant’s forces.
2. **Active behaviour**²: The robot complies with the participant’s forces until a perturbation occurs; only then does it instantly apply controlled forces in the opposite direction of imbalance to help the participant maintain balance.

To prevent exaggerated differences between conditions, we employed a between-subject study design. Participants were not informed in advance about the possible robot behaviors to quantify their perception and behavior towards unanticipated robot motions objectively. They are asked to perform a simple math calculation (running sum) as a secondary task while maintaining their balance on the regularly timed perturbed balance board. Each participant experienced 4 perturbations across 100 seconds while simultaneously performing the secondary task. We will henceforth refer to this 100-second duration

¹Participant experiencing [passive behaviour](#)

²Participant experiencing [active behaviour](#)

as the 'actual session'. Each participant had the opportunity to perform the primary task (experiences 2 perturbations) and the secondary task individually before performing them concurrently. Perturbations took place every 20 seconds. The secondary task contributes to the quantification of human mental load, as will be further detailed in Section 5.4.3, by simulating the real-world experience of maintaining balance while simultaneously engaging in another primary task, such as reading a sign while walking.

All our experiments received ethical approval from the University of Waterloo Human Research Ethics Board (protocol N. 45212) at the University of Waterloo. Before the experiment, participants received proper information and gave informed consent to participate in the study.

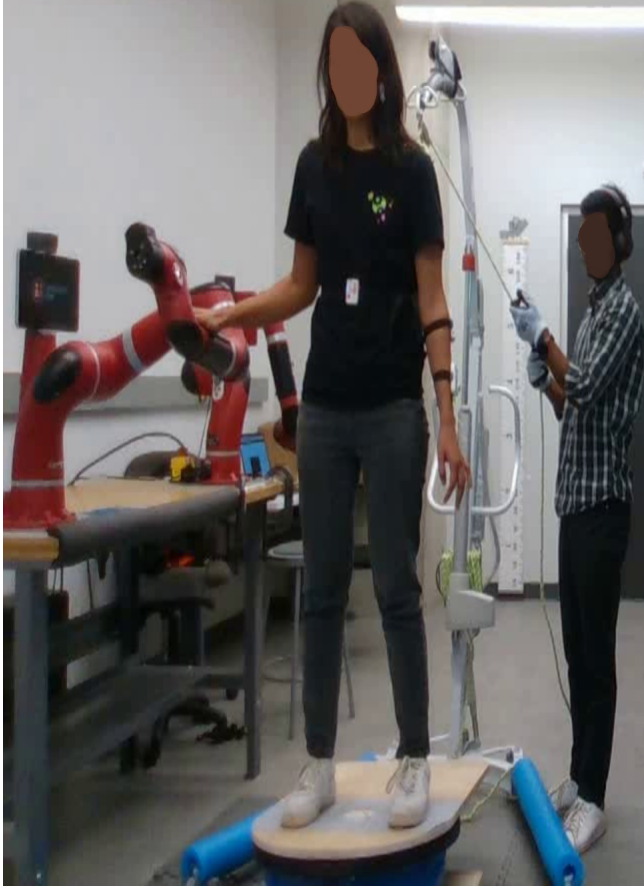
5.4.1 Equipment

We used the Rethink Robotics Sawyer collaborative serial manipulator, shown in Figure 5.1, which has 7 degrees of freedom and is equipped with joint torque sensors, to facilitate torque control mode implementation. PPG signals, to be converted into heart rate, were captured using an optical pulse sensor via Shimmer3 GSR+ with an earlobe clip. Skin temperature was monitored via a Shimmer3 Bridge Amplifier (BA)+ positioned on the distal phalanx of the fifth finger of the left hand. For primary task performance, we utilized IMUs from Shimmer3, placed at the participant's waist and at the center of the balance board. A 4.2 Kg sand kettlebell was employed as the soft sandbag, released using a pulley system. A half-ball balance board is used, with a 3mm thick plywood sheet attached to its surface. This plywood attachment guarantees safety by maintaining a safe distance between participants and the sandbag drop zone.

5.4.2 Robot Control

To implement the required behavior for each condition, we utilized low stiffness joint impedance control for passive behavior. Control parameters were determined through pilot experiments to enable the robot to operate within its stability margin with minimal stiffness in joint impedance control mode [162]. Uniform stiffness and damping values of 20 and 1, respectively, were applied to all joints.

For active behavior, we augmented joint impedance control with joint position control, providing an impression of applying opposing forces to aid participants in maintaining balance, as shown in the demonstration video of active behaviour. The robot controller switches from impedance control to position control when a perturbation occurs,



(a) Front view of the experiment setting



(b) Side view of the experiment setting

Figure 5.1: Interactive [pHRI](#) experiment setting.

The participant maintains their balance, prepared for the next perturbation, while a researcher wearing headphones and awaiting for the signal beep indicating the sandbag drop. The participant, while holding onto the robot for support, is equipped with an earlobe electrode for [PPG](#) signal measurement, an IMU around their waist, and a skin surface temperature probe on their fifth finger on the left hand. Another [Inertial Measurements Unit \(IMU\)](#) is placed on the centre of the balance board, right between the participant's feet.

and switches back to impedance control after 5 seconds, within which participants can balance. In position control mode, the robot sets its joint configurations to their current positions, except for the second joint from the base, which is incremented by 0.01 radians in the upward direction. This means that the robot remains in the low stiffness mode, akin to passive behavior, until a perturbation occurs. When a perturbation is triggered, the robot moves its second joint upward, becomes rigid for 5 seconds and then switches to the low stiffness mode again. Thus, synchronizing joint position control with perturbation occurrence was necessary. Parameters for joint position control were established via pilot experiments to ensure perceptible forces were applied by the robot to assist users in balance maintenance, addressing concerns raised in prior literature [76].

5.4.3 Measures

Based on our findings in Chapter 2, we opted to represent the human mental state by incorporating measurable dimensions within belief and cognitive ergonomics, including robot perception, mental workload, and stress.

This subsection begins by discussing the methodologies used to quantify human mental states to address **RQ1**. Following this, it outlines the approaches used to collect the necessary data to answer **RQ2**.

For reliable measures of the human mental state, we considered both objective and subjective approaches whenever feasible. Furthermore, we sought to understand the reasons behind the participants' responses to our quantitative measures by relying on open-ended questions.

Psychophysiology has correlated peripheral skin temperature with acute stress. According to Kim et al. [85], acute stress response of the **Sympathetic Nervous System (SNS)** is predominantly represented by an immediate response in the peripheral organs, such as heart rate and body temperature. Consequently, we chose to place the skin temperature probe on the distal phalanx of the fifth finger of the left hand as an objective measure of participant's stress levels. This decision was based on findings from Vinkers et al. [174], which showed that stress exposure resulted in changes in skin temperature following a gradient-like pattern, with decreases at distal skin locations such as the fingertip. Moreover, placing the skin temperature probe on the distal phalanx of the fifth finger is a common practice in **pHRI** [123]. It is noteworthy that Novak et al. [123] used skin temperature to indicate mental load rather than stress, diverging from the emphasis on stress in robotics and psychophysiology literature. This highlights the importance of Chapter 2 in addressing the inconsistent usage of human factor terminologies. In addition to assessing

skin temperature growth, we look forward to analyze the percentage difference between peripheral skin temperature during the actual session and its baseline. While previous studies in robotics literature have focused solely on average peripheral skin temperature [14], without comparing it to the baseline, this approach diverges from standard practices in psychology [174].

Furthermore, we opted to quantify stress by analyzing heart rate variation, employing PPG. Heart rate variations serve as physical biomarkers that humans perceive in response to stress [85].

Incorporating a secondary task alongside balance tasks to assess performance is a widespread practice in evaluating athletes' capabilities [149, 24]. The secondary task approach serves to gauge the residual mental load participants retain from the primary task. It is crucial to select a secondary task that does not interfere with or influence performance on the primary task. In other words, the secondary task should not draw participants' focus away from the primary task. Therefore, many studies examining mental load capacity while assessing athletes' balancing performance employ auditory and visual tasks involving recognition of consonants or vowels [142], simple mathematical calculations such as running sums [78], or countdowns [172]. Consistent with Hunter & Hoffman [78], we adopt an auditory running sum approach for the secondary task. In accordance with Hunter & Hoffman's approach [78], we selected numbers ranging from 1 to 6 as they are regarded as representing easy and medium difficulty levels in running sum tasks. Participants verbally provide responses and mentally retain the last calculated number to continue adding subsequent numbers. Numbers are verbally presented to participants upon response. A predetermined number sequence is repeated throughout the session. Participants are instructed that accuracy and speed are equally valued, requiring them to strike a balance between providing correct and prompt responses. To evaluate their performance in the secondary task, we devised the following performance index:

$$\begin{aligned}\text{Performance Index} &= 0.5(\text{Accuracy} + \text{Efficiency}) \\ \text{Accuracy} &= \frac{\text{Total number of correct answers}}{\text{Total number of attempted calculations}} \\ \text{Efficiency} &= \text{Total number of correct answers}\end{aligned}$$

It is important to note that we intentionally refrained from visually presenting the numbers to participants. We anticipated that their focus on regaining balance after each perturbation might compromise their ability to stand still and concentrate on the screen for clear visual perception. This decision was made recognizing that the overall performance of the secondary task could be influenced by the physical demands of the primary task, rather

than solely reflecting mental load. Thus, depending on a visual secondary task to assess participants' mental load induced by the primary task within our experimental setup may yield inaccurate results.

While questionnaires remain a prevalent method for quantifying mental states, the most commonly used questionnaire in robotics, the [NASA-TLX](#) [71] has shown major pitfalls [23]. Additionally, it encompasses irrelevant dimensions, such as physical demand, which deviates from cognitive load. Consequently, we have opted to adapt the [Students' Mental Load and Mental Effort in Biology Education-Questionnaire \(StuMMB-Q\)](#) [89] from educational science. This questionnaire evaluates both mental load and mental effort, and it has been validated in an educational context, particularly for biology students. Mental effort reflects an individual's invested cognitive capacity in a task, while mental load indicates the cognitive capacity required to process task complexity. The adapted version of this questionnaire is available in Appendix A.1.

According to our findings in Chapter 3, we employed [RoSAS](#) and perceived safety to subjectively evaluate robot perception.

As discussed in Section 5.2, the correlation between robot perception and peripheral skin temperature growth allows us to consider it as an objective measure for evaluating robot perception.

We further aimed to comprehend participants' reactions to our quantitative measures by incorporating open-ended questions and assessing their perceived usability of the system through the [System Usability Scale \(SUS\)](#) [151]. The open-ended questions encompassed: "Do you think the robot was helpful? Why or why not?", "What do you think the task of the robot was? Why or why not?", "What stood out to you the most about interacting with the robot?", and "What do you think of the interaction with the robot?" We prompted participants to elaborate on their responses to ensure thoroughness and clarity in their explanations.

In addition to capturing physiological signals and utilizing questionnaires to gauge participants' mental states, collecting data on participants' demographics, personality traits, and physical data is essential for addressing **RQ2**.

Demographic information, including factors such as age and gender, is crucial for understanding the composition of the participant sample. Personality traits are assessed using a widely recognized questionnaire, namely the [TIPI](#) [62].

We anticipate that certain physical data, such as distances between hands and feet and proximity to the robot, which have shown meaningful correlations with personality, interaction forces, and attitudes in active [pHRI](#) literature [75], may not result in meaningful

correlations in our case. This is because the nature of the task adopted for this experiment restricts participants’ movements to some extent.

However, to fully address **RQ2** and explore the potential of available physical data, we aim to quantify participants’ ability to maintain balance while utilizing the robot’s assistance. This entails participants wearing an IMU around their waist, with another IMU placed on the balance board.

5.4.4 System Design

This subsection delineates our endeavors to establish an optimal setting conducive to providing objective and reliable answers to our research questions. Our priorities include ensuring safety for both participants and researchers, maintaining a consistent experience across all participants, and ensuring system reliability despite variations in human behavior. By outlining these efforts, we justify the selection of our final setting. Towards the conclusion of this subsection, we provide a summary of the ultimate setting and detail the system’s control flow.

The setting for this study was highly complex as our aim was to evoke a sensation of imbalance without risking participants losing their balance or falling. This complexity included identifying an appropriate balance board, a suitable trigger behaviour for the robot’s unanticipated physical actions and a method for inducing perturbations. Furthermore, we conducted numerous pilot experiments to comprehend typical human reflexes that could prompt the robot to provide support when an imbalance occurred. Despite these efforts, relying on human reflexes to trigger the robot’s various behaviors proved to be unfeasible without diving into a complex control framework that is out of the scope of our study.

In our pursuit of developing a robust system capable of consistent performance across a range of human behaviors, we sought to implement a controlled imbalance mechanism, which included exploring wide range of balance boards and movable platforms. At first, we chose to employ the miniPRO Ninebot electric balancing vehicle from Segway. While accessing its functionalities posed challenges, notably through Bluetooth, our primary hurdle lay in determining an optimal margin—specifically, a leaning angle—for perturbation induction. This angle needed to elicit a sensation of imbalance without risking falls. Additionally, determining this angle proved highly sensitive to variables such as participants’ height and center of mass, which inherently differ between genders. Utilizing this approach would have mandated individual trial sessions to ascertain the appropriate lean angle for each participant, a process both time-intensive and potentially confounding due to its introduction of a learning parameter that could compromise the reliability of our results.

Consequently, we explored three alternative conventional balance boards, albeit lacking in control, before settling on the half-ball balance board, which best suited our study requirements. However, substantial effort and customization were still necessary to tailor it to our specific requirements.

We opted to manually release a soft sandbag onto the edge of the chosen balance platform to simulate the desired sensation of imbalance. This has required multiple pilot experiments to get to the suitable weight for different participant’s weights and heights without actually making them fall. Additionally, we customized the balance platform by affixing a thick plywood surface to increase its area, thereby facilitating consistent sandbag placement. This modification also provided a safety measure for experimenters by allowing them to drop the sandbag from a safe distance in the event of participant imbalance.

The height from which the sandbag was dropped was determined based on the minimum shoulder height of participants, ensuring it did not descend beyond their head or shoulders, as illustrated in Figure 5.1. A pulley system was utilized to drop the sandbag from a safe distance.

To signal the experimenter for sandbag release, headphones were provided, emitting a beep at the appropriate time. This approach accounted for the experimenter’s reflex time to the beep before dropping the bag, while also preventing the participant from knowing the exact timings of the droppings. This helped mitigating human errors.

After we considered an uncontrolled balance board, our objective was to trigger the joint position control in active conditions based on participants’ sensation of imbalance. This led us to explore methods for quantifying human imbalance. In existing neuroscience literature, imbalance is often quantified by estimating the human ankle torque threshold at which a person loses balance and falls [144, 116]. However, this method increases system complexity and further research in biomechanics that are out of the scope of this study. Another approach involves inferring potential human falls from body posture.

Initially, we investigated human pose estimation using MediaPipe. However, due to limitations in available graphic processing units, this method was unsuitable for real-time control and added unnecessary complexity. Additionally, attempts to infer potential human falls using IMUs on both the balance board and the participant’s waist were hindered by unreliable online streaming over Bluetooth, because of packet loss. This inconsistency in sensor data transmission raised concerns about achieving consistent robot behavior across participants.

As an alternative, we considered triggering the robot’s joint position control based on the velocity of the robot’s shoulder joint, as this joint is significantly impacted when participants attempt to maintain balance on the board. Triggering the robot’s behavior

using internal sensor data rather than external sensors offered promise, simulating the robot’s ability to make decisions based on its own sensor inputs. However, variations in human behavior, such as differences in height, weight, center of mass, and arm span, posed challenges to this approach.

Pilot experiments revealed that the velocity of the participant’s arm holding the robot did not consistently reach the threshold once a perturbation was induced. By the time the arm velocity reached the threshold, the participant’s body core had already deviated significantly from the center of the balance board, indicating a fall. Consequently, using arm joint velocity as a trigger for position control was deemed unsuitable.

Therefore, we opted for a time-based system instead of an event-based system to address our research questions, as approaches to target previously highlighted issues are out of the scope of the study.

Furthermore, while the main objective of the robot’s active behaviour is to switch from very low stiffness to very high stiffness whenever a perturbation occurred, the highest accomplished stiffness in the margin of stability was not sufficient for supporting of the participants and further increases of the stiffness would cause instability of the impedance controller. This has led us to switch to position control as a work around for the robot to exhibit the highest possible stiffness while being stable.

To achieve our goals, we determined that a time-based control system would provide reliable answers to our research questions with minimal complexity. Consequently, we developed modules written in Python within the [Robot Operating System \(ROS\)](#) framework: the behavior-based control module, perturbations module, Shimmers module, and trigger module.

The behavior-based control module handles activating the required condition (passive or active), switching between rigid and compliant behavior in the active condition, and recording the robot’s joint physical parameters (positions, velocity, accelerations, and torques).

The perturbations module signals the experimenter with a beep every 20 seconds.

The Shimmers module establishes Bluetooth communication with the Shimmer sensors and prepares them for recording.

Finally, the trigger module synchronizes all modules. It ensures that the dropping of the sandbag (i.e., the beeps) aligns with triggering the joint position mode in the active condition and with recording data from both the Shimmer sensors and the robot’s sensors.

5.4.5 Participants

Fifty-four participants were recruited from the University of Waterloo campus community, primarily comprising students. Only 35 participants were considered for the data analysis. Further explanation regarding discarded participants is provided towards the end of this subsection. The participants' ages ranged from 17 to 35 years ($\mu = 22.7, \sigma = 3.56$). The passive condition was randomly assigned 16 participants, while the active condition was randomly assigned 19 participants. Efforts were made to ensure age and gender balance within each condition. Additional details regarding the sample are provided in Table 5.1.

While our motivation for this study is supporting older adults independent living, it is important to acknowledge the lack of diversity in available physical assistive tasks involving unanticipated robot actions in a controlled environment. This justifies the reason behind studies adopting gaming scenarios for studying active behaviour [76, 75]. However, we opted to delve into unexplored area by investigating the active behavior of robots in assistive tasks with younger adults, with the trade-off of not exploring simpler sociable task with older adults, which have already been addressed extensively in literature [124, 60, 25, 52].

To ensure reliable answers for our research questions, we had a restricted eligibility criteria for weight, and height. Additionally, participants were required to have no history of heart or movement health issues, prefer using their right hand for physical support, and abstain from consuming caffeine, marijuana, or alcohol for at least 2 hours before the experiment.

Participants self-reported their balancing capability levels on a 10-point Likert scale (where 1 indicates low capability and 10 indicates high capability) and their experience with robots in general and specifically with the Sawyer robot on a 7-point Likert scale (where 1 indicates high experience and 7 indicates low experience). Both groups exhibited similar results with above-average balancing capability $M = 7$ and no prior experience with either Sawyer or robots in general $M = 5$.

Several participants were excluded from the data analysis due to various reasons. Three participants encountered technical issues during the experiments, while three others failed to meet the eligibility criteria. Additionally, 13 participants were unable to maintain their balance. Notably, among these 13 participants, 8 were males. Among the 54 participants initially recruited for the study, only 9 out of 17 males succeeded in maintaining their balance, resulting in a failure rate of 47%. In contrast, among the 37 females, 29 succeeded, leading to a failure rate of only 21.6%.

While the data acquired from the 16 participants failing to maintain balance on the balance board is not usable for the current study, it is nonetheless valuable for studying

Table 5.1: Interactive [pHRI](#) experiment sample description

	Passive	Active
No. of Participants	16	19
Female Participants	13	14
Male Participants	3	5
Age in years	21.6 ± 2.8	23.7 ± 3.8
Weight in Kg	59.2 ± 6.4	58.1 ± 7.3
Height in cm	167.1 ± 3.4	166.3 ± 3.7

humans falls and imbalance. Acquiring such data, depicting human poses during falling or loss of balance, is inherently challenging. We anticipate that this dataset will contribute significantly to other prominent research areas.

5.4.6 Procedure

Each participant visited the laboratory, dedicating approximately 45 minutes to complete the study. After obtaining participants’ informed consent, the experimenter addressed any ambiguity or confusion by providing verbal clarification for any questions raised. The experiment procedure consisted of four steps:

- Step #1: Preparation and Pre-trial Questionnaires
 - A screening phase was carried out to ensure participants meet our eligibility criteria.
 - Height and weight of participants were measured and recorded.
 - Participants completed pre-trial questionnaires, including demographic information (age, gender, ethnicity, profession, experience with robots, self-reported balancing capability), and the [TIPI](#).
 - Participants were equipped with a heart rate sensor, skin temperature sensor, an [IMU](#), and an audio recorder. This involved attaching an ear clip for [PPG](#) data collection, placing a skin temperature probe on the participant’s left distal phalanx, and positioning the [IMU](#) on the participant’s trunk (over their clothing), with the audio recorder secured to their shirt collar.

- For safety, participants were asked to wear suitable footwear and remove any accessories. They were then instructed to relax for 5 minutes to establish a baseline for PPG signals and skin temperature.
- Step #2: Instructions
 - Participants watched an instructional video³ demonstrating how to use the balance board, maintain balance, and perform the secondary task. Notably, the video did not include information about the robot’s capability of active or passive behavior.
 - A demonstration of the auditory secondary task was conducted with the participant to ensure comprehension.
 - Participant comfort was ensured by having them step over a chair of similar height to the balance board and hold onto the robot. This step helped determine a comfortable distance from the robot based on the participant’s height and arm span and gave the participant the chance to feel the robot’s stiffness for the first time.
- Step #3: The Experiment
 - Participants stood on the balance board, with the option to hold onto the robot, and attempted to maintain balance without perturbations.
 - Data recording commenced, including PPG signals, skin temperature, IMU data, and video recordings. Additionally, the data recording includes supplementary data aimed at providing a comprehensive understanding of the study, such as depth cameras capturing frontal and lateral views, as well as the robot’s log encompassing positions, velocities, accelerations, and torques of the joints and end effector. It is worth noting that the analysis of this supplementary information is beyond the current scope of the study.
 - Two perturbations, occurring at 20-second intervals, were induced using a soft sandbag dropped onto the balance board’s edge to familiarize participants with the task.
 - Six perturbations, occurring at 20-second intervals, were induced using the sandbag dropped onto the balance board’s edge. The first two perturbations served to familiarize participants with the task, with the secondary task commencing only after the third perturbation. Consequently, the actual session, where participants simultaneously performed the balancing task and the running sum,

³[Instructional Video](#)

lasted for 100 seconds, during which participants experienced four perturbations.

- After the actual session, participants were instructed to step down from the balance board, and sensors were removed and disinfected.
 - Participants completed post-trial questionnaires assessing robot perception, cognitive load, usability, and provided open-ended feedback.
- Step #4: Debriefing
 - The experimenter disclosed the primary objective of the experiment and addressed any participant questions.
 - Participants received remuneration and a debriefing letter.

5.4.7 Data Processing

Objective data analysis is conducted solely during the 100-second actual session. To calculate the performance index, video and audio recordings are thoroughly reviewed to assess the efficiency and accuracy of each participant’s responses. Additionally, temperature gradients are calculated based on linear trends observed in each participant’s temperature plot, with baseline and actual temperatures averaged to determine the percentage difference.

For PPG data, a low-pass filter is applied, and a cutoff frequency is determined individually for each participant to identify systolic and diastolic pulses during both baseline and actual sessions. However, a malfunction occurred with the Shimmer3 GSR+ sensor, rendering data from only 7 out of the 35 participants usable. Consequently, further analysis of PPG data was discontinued due to the insufficient sample size for statistical analysis.

Although the loss of PPG data would have provided an insightful inference of participants’ stress levels, we can still adequately address our research questions by relying on skin temperature measurements.

To assess participants’ ability to maintain balance on the balance board, we utilized the gyroscope data from the z-axis of the IMUs to calculate their imbalance durations. A low-pass filter was applied to the data, and a participant-specific threshold was determined based on their average balance maintenance. Participants were considered imbalanced if their values exceeded this threshold. At the beginning of the actual session, 4 seconds were excluded to eliminate the impact of the second perturbation induced for familiarity with the task. We considered excluding those 4 seconds based on pilot experiment durations without assistance.

Our adopted questionnaires to evaluate participants' subjective responses, we calculated scores for each measurable dimension by averaging individual items within each sub-scale.

5.5 Results

Figures 5.2 - 5.4 present box plots for each measurable dimension assessed. Each boxplot includes a median line, with the edges indicating the 25th and 75th percentiles. Whiskers extend up to 1.5 times the interquartile range, with outliers marked as 'o'.

As depicted in Figure 5.3, there was a decreasing trend in skin temperature growth across both groups, indicating that physical interaction with robots induces a certain level of stress, regardless of whether the interaction is active or passive. This is further supported by the calculated percentage difference in skin temperature, which showed a similar pattern across participants within the same group.

Participants exhibited a slightly higher performance index during interactions with the robot in the passive behavior condition compared to the active behavior condition, as illustrated in Figure 5.3.

The scores from the [StuMMB-Q](#), presented in Figure 5.4, indicate that mental load is significantly lower in the passive condition than in the active condition. However, the level of mental effort exerted across the two groups is somewhat comparable.

Both objective and subjective measures of cognitive ergonomics indicate that active behavior imposes a higher cognitive burden on participants compared to passive behavior.

Robot perception, chosen as a representation of belief in this study, is assessed through both subjective and objective measures.

As previously discussed regarding the stress objective measure, higher skin temperatures and lower skin temperature growths were observed in the passive behavior group compared to the active behavior group. This indicates that participants in the passive behavior condition perceived the robot more positively than those in the active behavior condition. However, confirmation or rejection of this interpretation will be explored further through correlation analysis between measurable dimensions, which will be discussed later in this section.

The subjective measure of robot perception, depicted in Figure 5.5, suggests that all relevant dimensions are comparable to some extent. Passive behavior tends to yield slightly lower scores in competence, warmth, and discomfort, while scoring similarly to active behavior in perceived safety.

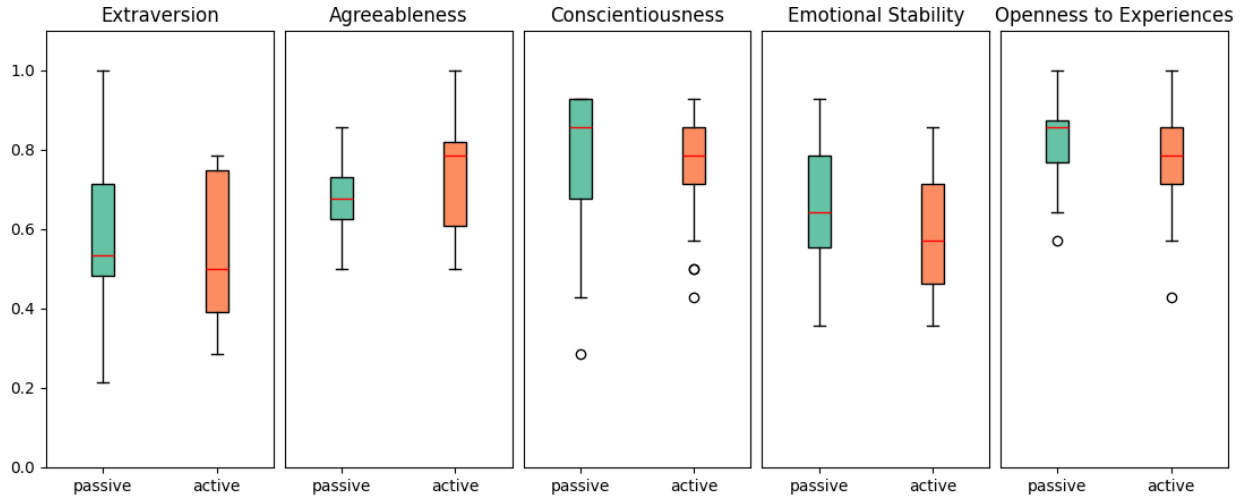


Figure 5.2: Boxplots illustrate participants' personality traits according to [TIPI](#). Each personality trait is normalized to fall within the range of 0 to 1.

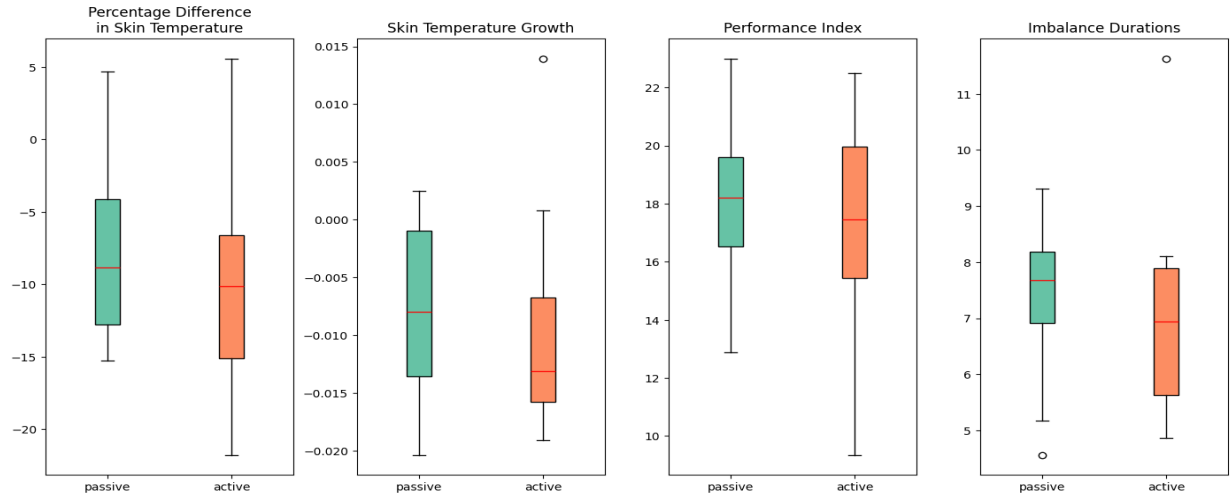


Figure 5.3: Boxplots depict participants' objective stress and robot perception measures, including skin temperature growth in $^{\circ}\text{C}/\text{s}$ and the percentage difference in skin temperature, alongside cognitive load measures represented by the performance index and overall performance indicated by the total imbalance durations in seconds across the actual session. Sign of percentage differences in skin temperature indicates whether there is an increase or decrease compared to baseline.

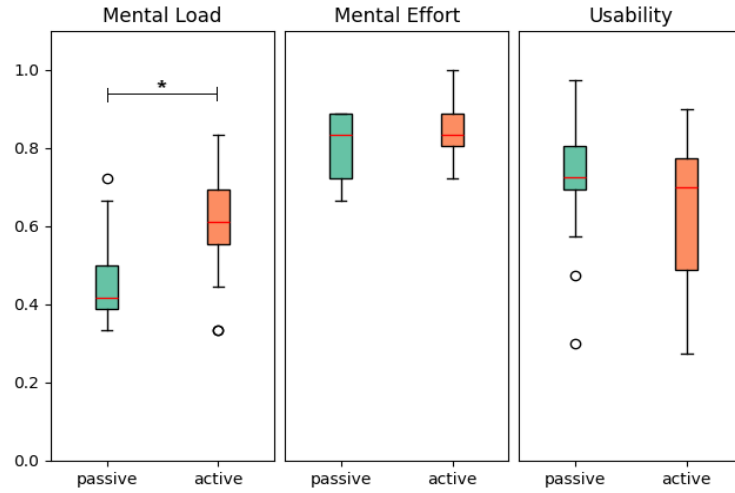


Figure 5.4: Boxplots illustrate participants' subjective cognitive load, encompassing mental load, and mental effort. In addition to usability ratings. Each measurable dimension ratings are normalized to fall within the range of 0 to 1. Significance levels ($*$:= p -value < 0.05) is indicated on lines between conditions.

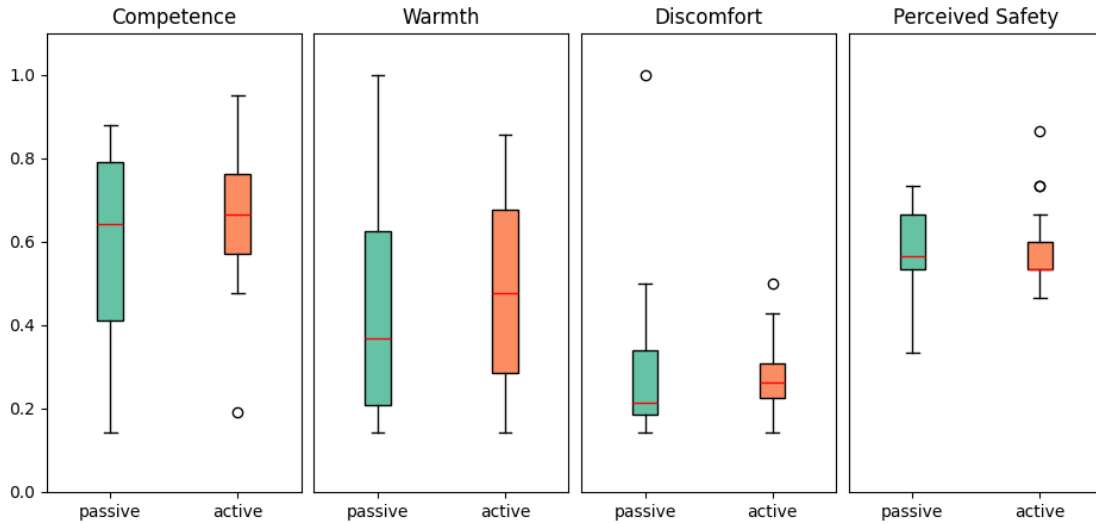


Figure 5.5: Boxplots depict participants' subjective perceptions of the robot, including competence, discomfort, warmth, and perceived safety. Each measurable dimension ratings are normalized to fall within the range of 0 to 1.

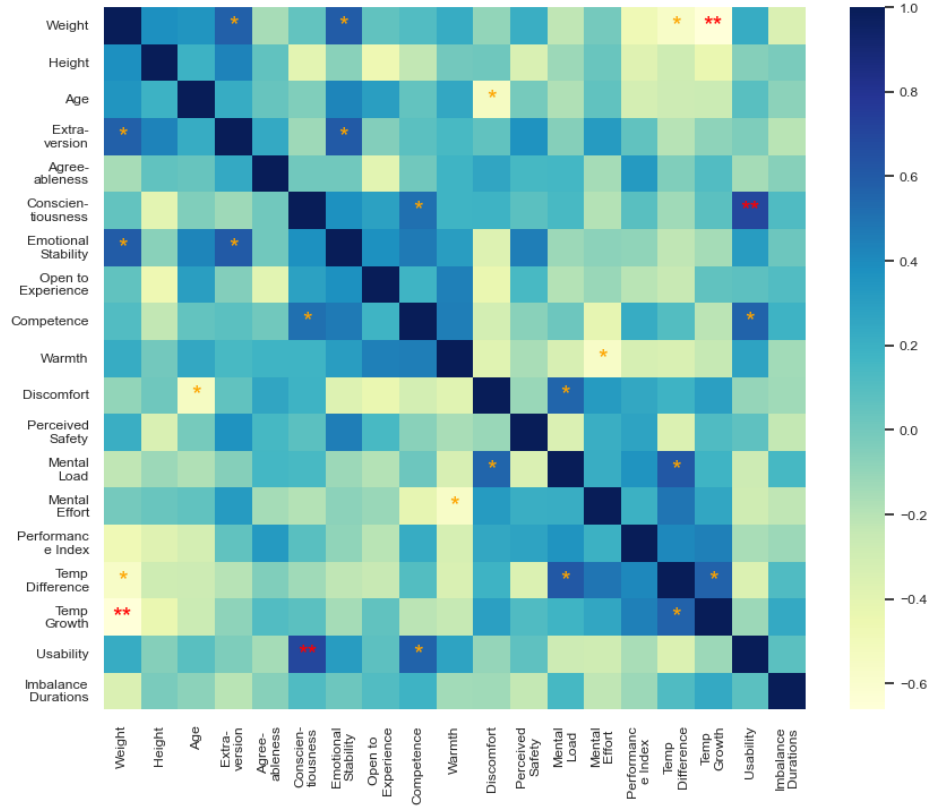


Figure 5.6: Heatmap illustrates Spearman's correlations observed for the **passive** condition between demographics, personality traits, physical data, physiological data, and measurable dimensions representing participants' mental states. Correlations marked with '*', and '**' indicate a p -value < 0.05 , and < 0.01 , respectively.

While the results indicate alignment between objective and subjective measures in cognitive ergonomics, belief objective and subjective measures typically do not exhibit such alignment.

The SUS scores, illustrated in Figure 5.4, indicate that perceived usability of the robot during active behavior is lower and with greater variation compared to passive behavior. However, participants experienced shorter imbalance durations while interacting with the robot in the active behavior, as depicted in Figure 5.3.

5.5.1 Variance tests

To analyze significant differences between the two robot behaviors, we utilized the Mann-Whitney U test. We chose this non-parametric test based on the type of data being compared (i.e., continuous). Given the smaller sample size in each condition ($n < 30$), a normality check was not feasible, prompting the selection of this test. To minimize the risk of false positives, we applied the Bonferroni Correction test. Additionally, effect size calculations were conducted for any significant differences identified.

Statistical analysis for the **TIPI** scores revealed no significant differences between the 2 groups, as depicted in Figure 5.2. This similarity enhances the validity of comparisons among the remaining dependent variables.

There was a statistically significant difference in mental load scores between the passive condition ($\mu = 1.375, \sigma = 0.346$) and the active condition ($\mu = 1.8, \sigma = 0.39$), with a $p - value = 0.0044$ and $W = -2.85$. This difference was associated with a substantial effect size of $r = 0.78$. No significant differences were observed for any other dimension.

5.5.2 Correlations

To answer **RQ2**, we employed Spearman’s rank correlation coefficient (ρ) to identify correlations within the collected data for each condition, as shown in Figures 5.6 and 5.7. Several strong correlations ($|\rho| > 0.6$) have been identified in both conditions.

In the passive condition, conscientiousness positively correlates with usability, while weight negatively correlates with skin temperature growth. Additionally, mental load and the percentage difference in skin temperature exhibit a positive correlation, whereas warmth and mental effort show a negative correlation.

In the active condition, agreeableness exhibits a negative correlation with mental load, while competence demonstrates positive correlations with warmth and usability. Furthermore, discomfort displays a negative correlation with perceived safety.

It is noteworthy that competence and usability exhibit a strong and statistically significant correlation ($\rho \approx 0.6, p - value < 0.05$) across both conditions.

5.5.3 Answers to RQs

In conclusion, our findings in response to **RQ1** suggest that the robot’s active behavior has a negative impact on participants’ cognitive ergonomics, particularly in terms of

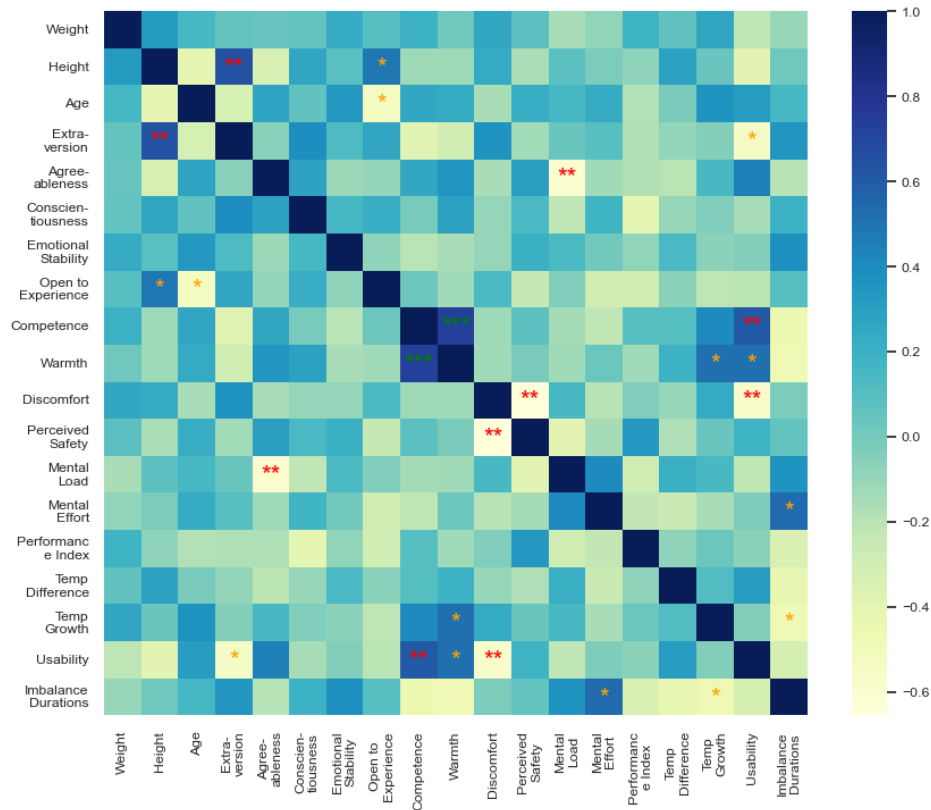


Figure 5.7: Heatmap illustrates Spearman's correlations observed for the **active** condition between demographics, personality traits, physical data, physiological data, and measurable dimensions representing participants' mental states. Correlations marked with '*', '**', and '***' indicate a $p - value < 0.05$, < 0.01 , and < 0.001 , respectively.

mental load. However, these behaviors do not seem to influence participants' perception of the robot. Regarding **RQ2**, cognitive ergonomics exhibited significant correlations among its measurable dimensions and pre-gathered data, such as weight and personality traits, among participants in the passive condition. Conversely, in the active condition, belief showed significant correlations among its measurable dimensions and a significant association emerged between pre-gathered data, such as personality traits, and cognitive ergonomics.

5.6 Discussion

Some limitations of the recruited sample negatively affect the generalizability of the study results. The participants recruited are predominantly university students, likely within a narrow age range, which limits the broader applicability of the findings. Moreover, the high failure rates in maintaining balance among males further diminish the generalizability of the results. The consistent challenges faced by males in maintaining balance suggest that females, with their lower center of mass, possess a significant advantage in this regard.

The results indicate that interacting with the robot in the passive behavior is more cognitively ergonomic compared to the active behavior, as evident across various measurable dimensions of cognitive ergonomics. On the other hand, belief measurable dimensions showed contrasting results. While the robot was perceived as more competent and comfortable in the active behavior compared to the passive behavior, it was also perceived as less friendly.

While the active behavior may have contributed to shorter imbalance durations compared to the passive behavior, suggesting a decrease in mental load during active behavior, the unexpected movement of the arm could have imposed additional cognitive burden on participants, increasing the cognitive capacity required to process the task complexity (mental load). In addition to the cognitive capacity occupied by overcoming perturbations and managing secondary tasks simultaneously, participants may have also been preoccupied with concerns about the movement's implications, such as its purpose. This is supported by some open-ended responses, where participants speculated about the reason behind the movement, with some perceiving it as adaptive and helpful, while others questioned its intent. This is further supported by the higher secondary task performance scores shown in the passive behaviour as opposed to the active behaviour. On the other hand, mental effort scores across both groups were comparable suggesting that, despite the robot's physical behaviour, participants perceived their exerted mental effort to perform the tasks similarly.

It is worth noting that these findings contradict those of [75], where it was observed that individuals with positive attitudes towards robots experienced lower cognitive load. This highlights the importance of considering the nature of the mutual task performed by the parties, such as gaming versus a balancing task.

The perception of the movement as helpful, as indicated by some open-ended responses, might have contributed to participants viewing the active behavior as competent, and less intimidating compared to the passive behavior.

The statistically significant difference observed only in mental load somewhat limits these conclusions. Nonetheless, the small sample size, from a statistical point of view, suggests a possibility of Type II error, indicating that recruiting a larger number of participants could potentially reveal more significant differences. Furthermore, a within-subject design could have been a more suitable approach for this study to exaggerate the differences between behaviors, hence making noticeable robot physical actions. Analysis of open-ended responses revealed that not all participants explicitly mentioned whether the robotic arm had moved up and become rigid after each perturbation. Therefore, it remains unclear whether all participants noticed this specific aspect of the robotic arm’s behavior.

While reliability assessment of questionnaires is not typically conducted in [HRI](#) studies, as discussed in Chapter 2, we have reservations about our conclusions drawn from the [StuMMB-Q](#) questionnaire. Therefore, we recognize the necessity of assessing its reliability before further building upon these results.

The statistically significant positive correlation between subjective measures (mental load and warmth) and objective measures (percentage difference in skin temperature and skin temperature growth, respectively) suggests that these objective measures have the potential to serve as reliable indicators for inferring belief and cognitive ergonomics dimensions during interaction. Additionally, the significant correlation between imbalance durations and cognitive ergonomics measures (mental effort and skin temperature growth) suggests that primary task performance can be utilized as an indicator of cognitive ergonomics.

The significant correlations between pre-gathered data (conscientiousness, agreeableness, age, and weight) and measurable dimensions (competence, mental load, discomfort, and temperature variations, respectively) present a promising supplementary approach for enhancing the inference of cognitive ergonomics and belief using objective measures.

However, the absence of the above correlations in both conditions introduces uncertainty, highlighting the need for further research in this area. Specifically, this inconsistency of correlations across conditions suggest that certain measures may be more appropriate for certain contexts even if both contexts belong to the broader scope of direct [pHRI](#). Nev-

ertheless, this suggestion may not hold true due to the small sample size, from a statistical point of view, which may have hindered the identification of consistent correlations across both conditions [153]. Furthermore, we caution against interpreting correlations between certain independent variables, such as weight, and physiological signals, such as temperature variations, solely from a psychological standpoint. The influence of such independent variables on physiological signals may be driven by human body physiology rather than solely the individual’s mental state. That said, performing clustering analysis may help mitigate this potential limitation.

The significant correlations observed between robot perception dimensions, such as competence and warmth, suggest that individuals who perceive the robot as competent are likely to also perceive it as friendly and warm. This relationship extends beyond direct pHRI to the broader domain of HRI, as noted by the authors of RoSAS [29].

Similarly, individuals who perceive high levels of safety during physical coupling with the robot tend to experience greater comfort during the interaction.

The significant correlations between belief dimensions (discomfort and warmth) and cognitive ergonomics measures (mental load and mental effort, respectively) align with the holistic conceptual model presented in Chapter 2, further emphasizing the cognitive relationship between them.

While correlations within pre-gathered data may have implications in other research fields, their analysis falls outside the scope of our current study.

The consistency observed within cognitive ergonomics dimensions; and between objective and subjective measures; as well as the demonstrated strong significant correlations, suggests the potential to infer cognitive load based on skin temperature and the performance of primary and secondary tasks. Conversely, the inconsistency demonstrated within belief dimensions; and between objective measures and some subjective measures, as well as the lack of correlations; highlights the need for further investigation in this area. Specifically, it raises the question of which objective measures are suitable for inferring attitudes towards the robot.

5.7 Conclusion

Our study aims to examine how different robot behaviors (active vs. passive pHRI) affect users’ mental states and contribute to the development of a human mental state model. We seek to correlate real-time measurable data with pre-gathered data and subjective responses to understand human mental states in pHRI during a physically assistive task.

Representing mental states through dimensions of cognitive ergonomics and belief, we formulated two research questions:

RQ1: Will an active robot behaviour impact the mental state of participants compared to passive robot behaviour in an assistive [pHRI](#) scenario?

Our findings suggest that passive robot interaction is generally more cognitively ergonomic than active interaction, despite nuanced subjective perceptions. While active behavior may reduce imbalance durations, it introduces additional cognitive burden due to unexpected arm movements.

We observed significant correlations between subjective and objective measures, indicating potential indicators for belief and cognitive ergonomics dimensions. However, certain correlations were absent in both conditions, necessitating further research. It is crucial to exercise caution when interpreting correlations between independent variables and physiological signals solely from a psychological standpoint.

RQ2: What correlations exist between participants' demographics, personality traits, physical data, physiological data, and their mental state during a physical assistive [pHRI](#) scenario?

Cognitive ergonomics exhibited significant correlations with measurable dimensions and pre-gathered data, such as weight and personality traits, in the passive condition. Conversely, in the active condition, belief showed significant correlations with measurable dimensions, and a significant association emerged between pre-gathered data, such as personality traits, and cognitive ergonomics.

In conclusion, our study highlights the importance of exploring cognitive ergonomics and belief in physical human-robot interaction contexts to develop human mental models. These models can enhance robots' decision-making capabilities, contributing to their effectiveness in various tasks.

Chapter 6

Conclusion

This thesis aims to investigate potential real-time quantification measures for human mental states in [pHRI](#). The primary aim is to aid in the development of a human mental state predictor, thereby endowing robots with advanced decision-making capabilities in [pHRI](#) scenarios.

To establish the research foundation, we conducted a literature review to explore approaches for quantifying human mental states. We aimed to address fundamental questions: 'What is human mental state?' and 'How can we quantify human mental state?'. Through this exploration, we identified cognitive ergonomics and belief as the two human factor types associated with cognitive aspects that best represent the human mental state in our study. These dimensions encompass robot perception for belief and mental workload and stress for cognitive ergonomics. While for quantification, a combination of objective measures (physical data and physiological signals) and subjective measures (questionnaires) were used.

This literature review contributed to the research community by identifying the most common quantifiable human factors in [pHRI](#), noting the factors typically addressed together, and determining the most commonly used assessment approaches. We also collected and classified proposed assessment approaches for each human factor and created an initial map of the common contexts and applications in [pHRI](#). We were able to provide an unbiased overview of the field and identify research gaps.

The proposed mapping scheme has revealed a demand for additional physical assistive applications in this domain, motivating us to adopt such an application into one of our user studies. Additionally, the studies have validated several anticipations made in the literature review. The prevalence of studies assessing different measurable dimensions

together, although not of the same human factor type, prompted us to investigate two human factor types simultaneously (cognitive ergonomics and belief), while also exploring the relationships between measurable dimensions within each type. Our findings indicate strong significant correlations between measurable dimensions across human factor types and within each type. Furthermore, these correlations align with the definitions adopted for each human factor type, providing further support for the suitability of our definitions. We believe that identifying and understanding these correlations can significantly contribute to the development of a robust human mental state predictor.

Furthermore, we aimed to address the gap identified in the literature review regarding the quantification of cognitive ergonomics within direct [pHRI](#) contexts through our balancing experiment. Our finding showed that lumping measurable dimensions into broader categories of human factor types holds promising for mitigating the diversity of terminologies used in the field. The effectiveness of this vision is highlighted by the consistency observed in cognitive ergonomics between its objective and subjective measures, as well as the robust correlations identified. These findings imply that cognitive ergonomics levels can be inferred based on factors such as skin temperature and the performance of primary and secondary tasks. In other words, the objective approaches investigated can be reasonably utilized to infer cognitive ergonomics levels without necessarily discerning differences between individual measurable dimensions within the human factor type. On the other hand, the inconsistency shown between belief dimensions, in addition to the inconsistency between the objective and some of the subjective measures, and the lack of correlations poses the need for further investigation in this area.

In alignment with our perspective that questionnaires serve as a preliminary assessment tool, and recognizing the necessity of directing efforts towards developing objective measures for human factor types (cognitive ergonomics and belief) insufficiently addressed in [pHRI](#), to facilitate the implementation of potential real-time measures for the development of a human mental state predictor, we focused on evaluating the reliability of existing questionnaires ([RoSAS](#) and perceived safety from Godspeed) within [pHRI](#). Additionally, we sought to adopt questionnaires from diverse fields ([StuMMB-Q](#)) to address significant gaps in [HRI](#). This approach is essential as assessing the reliability of subjective measures is crucial before identifying suitable objective measures. It is important to note that while competence and warmth from [RoSAS](#) were considered reliable subjective tools, additional efforts are needed to explore the suitability of discomfort and perceived safety. Moreover, assessing the reliability of [StuMMB-Q](#) is crucial to address a significant gap in [HRI](#), specifically the absence of a subjective tool for measuring cognitive load.

In response to our inquiry raised in the literature review, which questioned the generalizability of correlations between measurable dimensions and physical or physiological data

across diverse pHRI scenarios, we conducted the balance experiment. Our investigation aimed to ascertain whether the significance between objective and subjective measures persists across different pHRI scenarios. The findings revealed consistent correlations between skin temperature and positive attitudes within one of the measurable dimensions assessing robot perception. Specifically, a positive significant correlation was observed between warmth and temperature growth, which is consistent with the results reported by literature in social pHRI scenario [14].

When discussing correlations, it is crucial to acknowledge that while Spearman’s correlation is widely utilized in HRI for assessing relationships between objective and subjective measures, it primarily detects linear correlations, potentially missing other types of correlations. Therefore, exploring non-linear correlations within this data could unveil additional insights. However, we anticipate that machine learning models might offer more promise in regards of interpreting mental states, as even after identifying linear or non-linear correlations, establishing causal relationships requires rigorous investigation. Nonetheless, the challenge lies in the extensive training required for machine learning models on diverse datasets to ensure reliable results, which is particularly intimidating in pHRI due to the difficulty in collecting sufficiently large datasets.

Despite not being our primary objective, the exploration of impact of human factor types has brought the importance of task selection to our attention. The contradiction in cognitive load results between the literature utilizing gaming scenarios [75] and the balancing assistive task we adopted, despite both considering active pHRI scenarios, features the importance of carefully considering the nature of the collaborative task performed. This is further emphasized by the unexpected findings of the physical exercise experiment, where older adults perceived the robot displaying facial expressions as less competent compared to not displaying any, contrary to findings in social HRI literature.

Although our findings demonstrate a statistically significant correlation between subjective measures and objective measures, suggesting that these objective measures hold promise as reliable indicators for inferring cognitive ergonomics levels during interactions, we acknowledge that confirming their reliability necessitates establishing causal relationships. However, conducting such studies is challenging, as they necessitate scenarios where, for example, one group of participants experiences high stress levels induced by physical interaction with the robot, while another group experiences no or very low stress levels, with both groups having their skin temperature monitored. Only then, if significant differences are observed between the skin temperature of both groups, can skin temperature be deemed reliable for measuring stress levels in such a scenario.

Nevertheless, the significant correlations between pre-gathered data (baseline emotional states and gender) and measurable dimensions revealed in both user studies offer a promis-

ing supplementary approach for enhancing the inference of cognitive ergonomics and belief using objective measures. However, this requires the adoption of methods such as clustering analysis for more consolidated results.

We firmly believe that through the efforts outlined in this thesis, we have achieved the first fundamental milestones towards the development of a human mental state predictor. This includes establishing the foundation for potential real-time objective measures of cognitive ergonomics and belief, as well as creating a dataset suitable for training machine learning models.

6.1 Future Work

In research exploring potential real-time measures to quantify the human mental state in [pHRI](#), several key areas demand further exploration to enhance our understanding and development of effective interaction with robots. Current research highlights inconsistencies among belief dimensions and between objective and subjective measures, necessitating deeper investigation. Future studies could delve into psychophysiology to identify objective measures capable of accurately inferring belief levels or attitudes toward robots in real time. Additionally, longitudinal studies could work on confirming the effectiveness of cognitive ergonomics measures across various interactive scenarios.

Questionnaire assessment remains integral to [pHRI](#) research. While some dimensions assessing robot perception have demonstrated reliability, others, particularly perceived safety, require additional investigation to ensure their validity. Developing a reliable tool for quantifying subjective safety is crucial, given its pivotal role in [HRI](#). Additionally, assessing the reliability of existing subjective cognitive load measurement techniques from other fields to ascertain their suitability for our research domain could offer a viable solution to fill this gap in the literature. Alternatively, proposing new questionnaires tailored specifically to [pHRI](#) could provide another avenue for addressing this challenge.

Machine learning models present another promising area for further investigation into potential real-time measures. Developing models specifically tailored to [pHRI](#) has the potential to enhance the robustness and efficiency of predicting human mental states during interaction. Moreover, incorporating user specific information, such as baseline emotional states, gender, age, and personality traits, into these models could further enhance their interpretation capabilities.

In conclusion, addressing these areas of further research holds the potential to uncover promising real-time measures and pave the way for the development of more effective and

socially acceptable robotic systems capable of seamless integration into our communities. From examining real-time measures that accurately interpret human mental states to exploring suitable machine learning techniques that provide robot controllers with reliable information for advanced decision-making, there are numerous opportunities to enhance human lives through improved [pHRI](#) experiences.

References

- [1] Nourhan Abdulazeem and Yue Hu. “Human Factors Considerations for Quantifiable Human States in Physical Human-Robot Interaction”. en. In: *Sensors* (2023). DOI: [10.3390/s23177381](https://doi.org/10.3390/s23177381).
- [2] Nourhan Abdulazeem and Yue Hu. “The Impact of Robots’ Facial Emotional Expressions on Light Physical Exercises”. en. In: *Social Robotics, ICSR 2024*. Lecture Notes in Computer Science. Singapore: Springer Nature, 2024, pp. 216–229. DOI: [10.1007/978-981-99-8715-3_19](https://doi.org/10.1007/978-981-99-8715-3_19).
- [3] Global Coalition on Aging. *Building The Caregiving Workforce Our Aging World Needs*. July 2021. URL: https://globalcoalitiononaging.com/wp-content/uploads/2021/06/GCOA_HI_Building-the-Caregiving-Workforce-Our-Aging-World-Needs_REPORT-FINAL_July-2021.pdf.
- [4] B Alenljung et al. “User Experience of Conveying Emotions by Touch”. English. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Ed. by A Howard, K Suzuki, and L Zollo. IEEE, 2017, pp. 1240–1247.
- [5] Jacopo Aleotti, Vincenzo Micelli, and Stefano Caselli. “Comfortable robot to human object hand-over”. In: *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. ISSN: 1944-9437. Sept. 2012, pp. 771–776. DOI: [10.1109/ROMAN.2012.6343845](https://doi.org/10.1109/ROMAN.2012.6343845).
- [6] W. Amanhoud et al. “Contact-initiated shared control strategies for four-arm supernumerary manipulation with foot interfaces”. In: *International Journal of Robotics Research* 40.8-9 (2021), pp. 986–1014. DOI: [10.1177/02783649211017642](https://doi.org/10.1177/02783649211017642).
- [7] M Ammi et al. “Haptic Human-Robot Affective Interaction in a Handshaking Social Protocol”. English. In: *2015 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’15)*. IEEE, 2015, pp. 263–270. DOI: [10.1145/2696454.2696485](https://doi.org/10.1145/2696454.2696485).

- [8] R Andreasson et al. “Affective Touch in Human-Robot Interaction: Conveying Emotion to the Nao Robot”. English. In: *International Journal of Social Robotics* 10.4 (Sept. 2018), pp. 473–491. DOI: [10.1007/s12369-017-0446-3](https://doi.org/10.1007/s12369-017-0446-3).
- [9] Rachel H. Y. Au et al. “Robot Touch to Send Sympathy: Divergent Perspectives of Senders and Recipients”. en. In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Sapporo, Japan: IEEE, Mar. 2022, pp. 372–382. ISBN: 978-1-66540-731-1. DOI: [10.1109/HRI53351.2022.9889419](https://doi.org/10.1109/HRI53351.2022.9889419).
- [10] Armin Avaei et al. “An Incremental Inverse Reinforcement Learning Approach for Motion Planning with Separated Path and Velocity Preferences”. en. In: *Robotics* 12.2 (Apr. 2023). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 61. DOI: [10.3390/robotics12020061](https://doi.org/10.3390/robotics12020061).
- [11] J Avelino et al. “The Power of a Hand-shake in Human-Robot Interactions”. English. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1864–1869.
- [12] Laura Aymerich-Franch and Iliana Ferrer. “Liaison, safeguard, and well-being: Analyzing the role of social robots during the COVID-19 pandemic”. In: *Technology in society* 70 (2022), p. 101993.
- [13] Uba Backonja et al. “Comfort and attitudes towards robots among young, middle-aged, and older adults: a cross-sectional study”. In: *Journal of Nursing Scholarship* 50.6 (2018), pp. 623–633.
- [14] WA Bainbridge et al. “A Methodological Outline and Utility Assessment of Sensor-based Biosignal Measurement in Human-Robot Interaction A System for Determining Correlations Between Robot Sensor Data and Subjective Human Data in HRI”. English. In: *International Journal of Social Robotics* 4.3 (Aug. 2012), pp. 303–316. DOI: [10.1007/s12369-012-0146-y](https://doi.org/10.1007/s12369-012-0146-y).
- [15] Christoph Bartneck et al. *Human-Robot Interaction*. en. 1st ed. Cambridge University Press, Feb. 2020.
- [16] Christoph Bartneck et al. “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots”. en. In: *International Journal of Social Robotics* 1.1 (Jan. 2009), pp. 71–81.
- [17] R. Behrens et al. “A Statistical Model to Determine Biomechanical Limits for Physically Safe Interactions With Collaborative Robots”. In: *Frontiers in Robotics and AI* 8 (2022). DOI: [10.3389/frobt.2021.667818](https://doi.org/10.3389/frobt.2021.667818).

- [18] Ashwin Sadananda Bhat et al. “Plant robot for at-home behavioral activation therapy reminders to young adults with depression”. In: *ACM Transactions on Human-Robot Interaction (THRI)* 10.3 (2021), pp. 1–21.
- [19] Achin Bhowmik. *Forbes Council Post: The Future Of Multifunctional Devices In Our Everyday Lives*. en. Jan. 2024. URL: <https://www.forbes.com/sites/forbestechcouncil/2024/01/22/the-future-of-multifunctional-devices-in-our-everyday-lives/>.
- [20] A.E. Block and K.J. Kuchenbecker. “Emotionally Supporting Humans Through Robot Hugs”. In: *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2018, pp. 293–294.
- [21] Alexis E Block et al. “In the Arms of a Robot: Designing Autonomous Hugging Robots with Intra-Hug Gestures”. en. In: *ACM Transaction on Human-Robot Interaction* 12.2 (2022), p. 47.
- [22] N. Bock, L. Hoffmann, and R.-V.D.A. Pütten. “Your Touch Leaves Me Cold, Robot.” In: *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2018, pp. 71–72. DOI: [10.1145/3173386.3177036](https://doi.org/10.1145/3173386.3177036).
- [23] Per Øivind Braarud. “Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human-machine work”. en. In: *International Journal of Industrial Ergonomics* 86 (Nov. 2021), p. 103233. DOI: [10.1016/j.ergon.2021.103233](https://doi.org/10.1016/j.ergon.2021.103233).
- [24] Steven Broglio, Phillip Tomporowski, and Michael Ferrara. “Balance Performance with a Cognitive Task: A Dual-Task Testing Paradigm”. In: *Medicine and science in sports and exercise* 37 (Apr. 2005), pp. 689–95. DOI: [10.1249/01.MSS.0000159019.14919.09](https://doi.org/10.1249/01.MSS.0000159019.14919.09).
- [25] RB Burns et al. “Endowing a NAO Robot With Practical Social-Touch Perception”. English. In: *Frontiers in Robotics and AI* 9 (Apr. 2022). DOI: [10.3389/frobt.2022.840335](https://doi.org/10.3389/frobt.2022.840335).
- [26] Baptiste Busch, Marc Toussaint, and Manuel Lopes. “Planning Ergonomic Sequences of Actions in Human-Robot Interaction”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. ISSN: 2577-087X. May 2018, pp. 1916–1923. DOI: [10.1109/ICRA.2018.8462927](https://doi.org/10.1109/ICRA.2018.8462927).
- [27] Baptiste Busch et al. “Postural optimization for an ergonomic human-robot interaction”. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. ISSN: 2153-0866. Sept. 2017, pp. 2778–2785. DOI: [10.1109/IROS.2017.8206107](https://doi.org/10.1109/IROS.2017.8206107).

- [28] M.G. Carmichael, S. Aldini, and D. Liu. “Human user impressions of damping methods for singularity handling in human-robot collaboration”. In: *Australasian Conference on Robotics and Automation, ACRA*. Vol. 2017-December. Semantic Scholar, 2017, pp. 107–113.
- [29] Colleen M. Carpinella et al. “The Robotic Social Attributes Scale (RoSAS)”. In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2017, pp. 254–262.
- [30] Chien-Hsiung Chen and Xiaoyu Jia. “Effects of Head Shape, Facial Features, Camera, and Gender on the Perceptions of Rendered Robot Faces”. en. In: *International Journal of Social Robotics* 15.1 (Jan. 2023), pp. 71–84.
- [31] L. Chen, L.F.C. Figueredo, and M.R. Dogar. “Planning for Muscular and Peripersonal-Space Comfort during Human-Robot Forceful Collaboration”. In: *IEEE-RAS International Conference on Humanoid Robots*. Vol. 2018-November. IEEE, 2019, pp. 409–416. DOI: [10.1109/HUMANOIDS.2018.8624978](https://doi.org/10.1109/HUMANOIDS.2018.8624978).
- [32] T.L. Chen and C.C. Kemp. “Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots”. In: *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*. IEEE, 2010, pp. 367–374. DOI: [10.1145/1734454.1734579](https://doi.org/10.1145/1734454.1734579).
- [33] T.L. Chen et al. “An Investigation of Responses to Robot-Initiated Touch in a Nursing Context”. In: *International Journal of Social Robotics* 6.1 (2014), pp. 141–161. DOI: [10.1007/s12369-013-0215-x](https://doi.org/10.1007/s12369-013-0215-x).
- [34] Tiffany L. Chen et al. “Touched by a robot: An investigation of subjective responses to robot-initiated touch”. In: *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ISSN: 2167-2148. Mar. 2011, pp. 457–464. DOI: [10.1145/1957656.1957818](https://doi.org/10.1145/1957656.1957818).
- [35] H. Claire et al. “Using expectancy violations theory to understand robot touch interpretation”. In: *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2020, pp. 163–165. DOI: [10.1145/3371382.3378314](https://doi.org/10.1145/3371382.3378314).
- [36] Enrique Coronado et al. “Evaluating quality in human-robot interaction”. en. In: *Journal of Manufacturing Systems* 63 (Apr. 2022), pp. 392–410.
- [37] H Cramer et al. “‘Give me a hug’: the effects of touch and autonomy on people’s responses to embodied social agents”. English. In: *Computer Animation and Virtual Worlds* 20.2-3 (June 2009), pp. 437–445. ISSN: 1546-4261. DOI: [10.1002/cav.317](https://doi.org/10.1002/cav.317).

- [38] Frédéric Dehais et al. “Physiological and subjective evaluation of a human–robot object hand-over task”. en. In: *Applied Ergonomics* 42.6 (Nov. 2011), pp. 785–791. DOI: [10.1016/j.apergo.2010.12.005](https://doi.org/10.1016/j.apergo.2010.12.005).
- [39] L.V. Der Spaa et al. “Predicting and Optimizing Ergonomics in Physical Human-Robot Cooperation Tasks”. In: *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 1799–1805. DOI: [10.1109/ICRA40945.2020.9197296](https://doi.org/10.1109/ICRA40945.2020.9197296).
- [40] Yadolah Dodge. *The Concise Encyclopedia of Statistics*. English. 2010th edition. Philadelphia, Pa.: Springer, Apr. 2008. ISBN: 978-0-397-51837-1.
- [41] Iveta Eimontaite et al. “Language-free graphical signage improves human performance and reduces anxiety when working collaboratively with robots”. en. In: *The International Journal of Advanced Manufacturing Technology* 100.1-4 (Jan. 2019), pp. 55–73. DOI: [10.1007/s00170-018-2625-2](https://doi.org/10.1007/s00170-018-2625-2).
- [42] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the human face*. Emotion in the human face: Guidelines for research and an integration of findings. Pages: xii, 191. Oxford, England: Pergamon Press, 1972.
- [43] V Evers et al. “The Evaluation of Empathy, Autonomy and Touch to Inform the Design of an Environmental Monitoring Robot”. English. In: *Social Robotics, ICSR 2010*. Ed. by SS Ge et al. Vol. 6414. Springer Nature, 2010, pp. 285–294.
- [44] A.T. Eyam, W.M. Mohammed, and J.L. Martinez Lastra. “Emotion-driven analysis and control of human-robot interactions in collaborative applications”. In: *Sensors* 21.14 (2021). DOI: [10.3390/s21144626](https://doi.org/10.3390/s21144626). URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85109025481&doi=10.3390%2fs21144626&partnerID=40&md5=32533a6ae5f139c5613bbf797ba8417c>.
- [45] Ronit Feingold Polak et al. “Differences between young and old users when interacting with a humanoid robot: a qualitative usability study”. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 2018, pp. 107–108.
- [46] F Ferland et al. “Taking Your Robot For a Walk: Force-Guiding a Mobile Robot Using Compliant Arms”. English. In: *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2013)*. Ed. by H Kuzuoka et al. IEEE, 2013, pp. 309–316.
- [47] F. Ficuciello, L. Villani, and B. Siciliano. “Variable Impedance Control of Redundant Manipulators for Intuitive Human-Robot Physical Interaction”. In: *IEEE Transactions on Robotics* 31.4 (2015), pp. 850–863. DOI: [10.1109/TR0.2015.2430053](https://doi.org/10.1109/TR0.2015.2430053).

- [48] L.F.C. Figueredo et al. “Human Comfortability: Integrating Ergonomics and Muscular-Informed Metrics for Manipulability Analysis during Human-Robot Collaboration”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 351–358. DOI: [10.1109/LRA.2020.3043173](https://doi.org/10.1109/LRA.2020.3043173).
- [49] N.T. Fitter and K.J. Kuchenbecker. “How Does It Feel to Clap Hands with a Robot?” In: *International Journal of Social Robotics* 12.1 (2020), pp. 113–127. DOI: [10.1007/s12369-019-00542-x](https://doi.org/10.1007/s12369-019-00542-x).
- [50] N.T. Fitter and K.J. Kuchenbecker. “Synchronicity Trumps Mischief in Rhythmic Human-Robot Social-Physical Interaction”. In: *Springer Proceedings in Advanced Robotics* 10 (2020), pp. 269–284. DOI: [10.1007/978-3-030-28619-4_24](https://doi.org/10.1007/978-3-030-28619-4_24).
- [51] N.T. Fitter and K.J. Kuchenbecker. “Teaching a robot bimanual hand-clapping games via wrist-worn IMUs”. In: *Frontiers Robotics AI* 5.JUN (2018). DOI: [10.3389/frobt.2018.00085](https://doi.org/10.3389/frobt.2018.00085).
- [52] Naomi Fitter et al. “Exercising with Baxter: Preliminary support for assistive social-physical human-robot interaction”. In: *Journal of NeuroEngineering and Rehabilitation* 17 (Feb. 2020). DOI: [10.1186/s12984-020-0642-5](https://doi.org/10.1186/s12984-020-0642-5).
- [53] Naomi T. Fitter and Katherine J. Kuchenbecker. “Designing and Assessing Expressive Open-Source Faces for the Baxter Robot”. en. In: *Social Robotics, ICSR 2016*. Vol. 9979. Cham: Springer International Publishing, 2016, pp. 340–350.
- [54] Federico Fraboni et al. “Human-Robot Collaboration During Assembly Tasks: The Cognitive Effects of Collaborative Assembly Workstation Features”. en. In: *21st Congress of the International Ergonomics Association (IEA 2021)*. Ed. by Nancy L. Black, W. Patrick Neumann, and Ian Noy. Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2022, pp. 242–249. DOI: [10.1007/978-3-030-74614-8_29](https://doi.org/10.1007/978-3-030-74614-8_29).
- [55] Marina Fujita, Ryu Kato, and Arai Tamio. “Assessment of operators’ mental strain induced by hand-over motion of industrial robot manipulator”. In: *19th International Symposium in Robot and Human Interactive Communication*. Sept. 2010, pp. 361–366. DOI: [10.1109/ROMAN.2010.5598689](https://doi.org/10.1109/ROMAN.2010.5598689).
- [56] M Gaitan-Padilla et al. “Physical Human-Robot Interaction Through Hugs with CASTOR Robot”. English. In: *Social Robotics, ICSR 2021*. Ed. by H Li et al. Vol. 13086. Springer Nature, 2021, pp. 814–818. DOI: [10.1007/978-3-030-90525-5_77](https://doi.org/10.1007/978-3-030-90525-5_77).
- [57] Y. Gao, H.J. Chang, and Y. Demiris. “User Modelling Using Multimodal Information for Personalised Dressing Assistance”. In: *IEEE Access* 8 (2020), pp. 45700–45714. DOI: [10.1109/ACCESS.2020.2978207](https://doi.org/10.1109/ACCESS.2020.2978207).

- [58] Cristina Getson and Goldie Nejat. “Socially assistive robots helping older adults through the pandemic and life after COVID-19”. In: *Robotics* 10.3 (2021), p. 106.
- [59] Moojan Ghafurian, Colin Ellard, and Kerstin Dautenhahn. “Social companion robots to reduce isolation: A perception change due to COVID-19”. In: *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II* 18. Springer. 2021, pp. 43–63.
- [60] I Giorgi et al. “Friendly But Faulty: A Pilot Study on the Perceived Trust of Older Adults in a Social Robot”. English. In: *IEEE Access* 10 (2022), pp. 92084–92096. DOI: [10.1109/ACCESS.2022.3202942](https://doi.org/10.1109/ACCESS.2022.3202942).
- [61] S Gopinathan, SK Otting, and JJ Steil. “A User Study on Personalized Stiffness Control and Task Specificity in Physical Human-Robot Interaction”. English. In: *Frontiers in Robotics and AI* 4 (Nov. 2017). DOI: [10.3389/frobt.2017.00058](https://doi.org/10.3389/frobt.2017.00058).
- [62] Samuel D Gosling, Peter J Rentfrow, and William B Swann. “A very brief measure of the Big-Five personality domains”. en. In: *Journal of Research in Personality* 37.6 (Dec. 2003), pp. 504–528. DOI: [10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1).
- [63] Luca Gualtieri et al. “Development and evaluation of design guidelines for cognitive ergonomics in human-robot collaborative assembly systems”. en. In: *Applied Ergonomics* 104 (Oct. 2022), p. 103807. DOI: [10.1016/j.apergo.2022.103807](https://doi.org/10.1016/j.apergo.2022.103807).
- [64] Arzu Guneyusu Ozgur et al. “Gamified motor training with tangible robots in older adults: a feasibility study and comparison with the young”. In: *Frontiers in aging neuroscience* 12 (2020), p. 59.
- [65] Kasper Hald, Matthias Rehm, and Thomas B. Moeslund. “Proposing Human-Robot Trust Assessment Through Tracking Physical Apprehension Signals in Close-Proximity Human-Robot Collaboration”. In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. Oct. 2019, pp. 1–6. DOI: [10.1109/RO-MAN46459.2019.8956335](https://doi.org/10.1109/RO-MAN46459.2019.8956335).
- [66] Adriana Hamacher et al. “Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction”. In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. Aug. 2016, pp. 493–500. DOI: [10.1109/ROMAN.2016.7745163](https://doi.org/10.1109/ROMAN.2016.7745163).
- [67] Peter A. Hancock, Aaron A. Pepe, and Lauren L. Murphy. “Hedonomics: The Power of Positive and Pleasurable Ergonomics”. In: *Ergonomics in Design* 13.1 (Jan. 2005). Publisher: SAGE Publications Inc, pp. 8–14. ISSN: 1064-8046. DOI: [10.1177/106480460501300104](https://doi.org/10.1177/106480460501300104).

- [68] K.S. Haring et al. “Touching an Android robot: Would you do it and how?” In: *Control, Automation, and Robotics*. 2015, pp. 8–13. DOI: [10.1109/ICCAR.2015.7165993](https://doi.org/10.1109/ICCAR.2015.7165993).
- [69] A Hayashi, LK Rincon-Ardila, and G Venture. “Improving HRI with Force Sensing”. English. In: *Machines* 10.1 (Jan. 2022). DOI: [10.3390/machines10010015](https://doi.org/10.3390/machines10010015).
- [70] Kerry He et al. “On-The-Go Robot-to-Human Handovers with a Mobile Manipulator”. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. Aug. 2022, pp. 729–734. DOI: [10.1109/RO-MAN53752.2022.9900642](https://doi.org/10.1109/RO-MAN53752.2022.9900642).
- [71] Jamison Heard, Caroline E. Harriott, and Julie A. Adams. “A survey of workload assessment algorithms”. In: *IEEE Transactions on Human-Machine Systems* 48.5 (Oct. 2018). Conference Name: IEEE Transactions on Human-Machine Systems, pp. 434–451. DOI: [10.1109/THMS.2017.2782483](https://doi.org/10.1109/THMS.2017.2782483).
- [72] Evelien Heyselaar. “The CASA theory no longer applies to desktop computers”. In: *Scientific Reports* 13.1 (2023), p. 19693.
- [73] T Hirano et al. “How Do Communication Cues Change Impressions of Human-Robot Touch Interaction?” English. In: *International Journal of Social Robotics* 10.1 (Jan. 2018), pp. 21–31. DOI: [10.1007/s12369-017-0425-8](https://doi.org/10.1007/s12369-017-0425-8).
- [74] Sarah Hopko, Jingkun Wang, and Ranjana Mehta. “Human Factors Considerations and Metrics in Shared Space Human-Robot Collaboration: A Systematic Review”. In: *Frontiers in Robotics and AI* 9 (2022).
- [75] Y. Hu et al. “Toward Active Physical Human–Robot Interaction”. en. In: *IEEE Transactions on Human Machine Systems* (2022).
- [76] Yue Hu et al. “Interact With Me: An Exploratory Study on Interaction Factors for Active Physical Human-Robot Interaction”. In: *IEEE Robotics and Automation Letters* 5.4 (Oct. 2020). Conference Name: IEEE Robotics and Automation Letters, pp. 6764–6771. ISSN: 2377-3766. DOI: [10.1109/LRA.2020.3017475](https://doi.org/10.1109/LRA.2020.3017475).
- [77] M Huber et al. “Evaluation of a novel biologically inspired Trajectory Generator in Human-Robot Interaction”. English. In: *RO-MAN 2009: The 18th IEEE International Symposium on Robot and Human Interactive Communication, Vols 1 and 2*. IEEE, 2009, pp. 712–+.
- [78] Megan C. Hunter and Mark A. Hoffman. “Postural control: visual and cognitive manipulations”. en. In: *Gait & Posture* 13.1 (Feb. 2001), pp. 41–48. DOI: [10.1016/S0966-6362\(00\)00089-8](https://doi.org/10.1016/S0966-6362(00)00089-8).

- [79] Sumio Ishii. “Meal-assistance Robot “My Spoon””. en. In: *Journal of the Robotics Society of Japan* 21.4 (May 2003). Publisher: The Robotics Society of Japan, pp. 378–381. DOI: [10.7210/jrsj.21.378](https://doi.org/10.7210/jrsj.21.378).
- [80] ISO. *ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. URL: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en> (visited on 04/23/2023).
- [81] Government of Japan. *Society 5.0*. en. URL: https://www8.cao.go.jp/cstp/english/society5_0/index.html.
- [82] M. Jindai and T. Watanabe. “Development of a handshake request motion model based on analysis of handshake motion between humans”. In: *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*. IEEE, 2011, pp. 560–565. DOI: [10.1109/AIM.2011.6026975](https://doi.org/10.1109/AIM.2011.6026975).
- [83] M. Jindai et al. “Handshake request motion model with an approaching human for a handshake robot system”. In: *2015 7th IEEE International Conference on Cybernetics and Intelligent Systems, CIS 2015 and Robotics, Automation and Mechatronics, RAM 2015*. IEEE, 2015, pp. 265–270. DOI: [10.1109/ICCIS.2015.7274632](https://doi.org/10.1109/ICCIS.2015.7274632).
- [84] J. Jørgensen, K.B. Bojesen, and E. Jochum. “Is a Soft Robot More “Natural”? Exploring the Perception of Soft Robotics in Human–Robot Interaction”. In: *International Journal of Social Robotics* 14.1 (2022), pp. 95–113. DOI: [10.1007/s12369-021-00761-1](https://doi.org/10.1007/s12369-021-00761-1).
- [85] Heena Kim et al. “Recent Advances in Multiplexed Wearable Sensor Platforms for Real-Time Monitoring Lifetime Stress: A Review”. en. In: *Biosensors* 13.4 (Apr. 2023). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 470. DOI: [10.3390/bios13040470](https://doi.org/10.3390/bios13040470).
- [86] W. Kim et al. “Anticipatory Robot Assistance for the Prevention of Human Static Joint Overloading in Human-Robot Collaboration”. In: *IEEE Robotics and Automation Letters* 3.1 (2018), pp. 68–75. DOI: [10.1109/LRA.2017.2729666](https://doi.org/10.1109/LRA.2017.2729666).
- [87] Barbara Kitchenham and Pearl Brereton. “A systematic review of systematic review process research in software engineering”. en. In: *Information and Software Technology* 55.12 (Dec. 2013), pp. 2049–2075. DOI: [10.1016/j.infsof.2013.07.010](https://doi.org/10.1016/j.infsof.2013.07.010).
- [88] Christian Kowalski et al. “Comparison of a VR-based and a rule-based robot control method for assistance in a physical human-robot collaboration scenario”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 722–728. DOI: [10.1109/RO-MAN53752.2022.9900779](https://doi.org/10.1109/RO-MAN53752.2022.9900779).

- [89] Moritz Krell. “Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence”. In: *Cogent Education* 4 (Jan. 2017). DOI: [10.1080/2331186X.2017.1280256](https://doi.org/10.1080/2331186X.2017.1280256).
- [90] H. Kunimura et al. *Baby robot "YOTARO"*. Vol. 6243 LNCS. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Pages: 8. International Conference on Entertainment Computing, ICEC 2010, 2010. DOI: [10.1007/978-3-642-15399-0_1](https://doi.org/10.1007/978-3-642-15399-0_1).
- [91] Laura Kunold. “Seeing is not Feeling the Touch from a Robot”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 1562–1569. DOI: [10.1109/RO-MAN53752.2022.9900788](https://doi.org/10.1109/RO-MAN53752.2022.9900788).
- [92] Marta Lagomarsino et al. “Pick the Right Co-Worker: Online Assessment of Cognitive Ergonomics in Human-Robot Collaborative Assembly”. In: *IEEE Transactions on Cognitive and Developmental Systems* (2022). Conference Name: IEEE Transactions on Cognitive and Developmental Systems, pp. 1–1. DOI: [10.1109/TCDS.2022.3182811](https://doi.org/10.1109/TCDS.2022.3182811).
- [93] Jens Lambrecht and Sebastian Nimpsch. “Human Prediction for the Natural Instruction of Handovers in Human Robot Collaboration”. In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. Oct. 2019, pp. 1–6. DOI: [10.1109/RO-MAN46459.2019.8956379](https://doi.org/10.1109/RO-MAN46459.2019.8956379).
- [94] Allison Langer et al. “Trust in socially assistive robots: Considerations for use in rehabilitation”. en. In: *Neuroscience & Biobehavioral Reviews* 104 (Sept. 2019), pp. 231–239. DOI: [10.1016/j.neubiorev.2019.07.014](https://doi.org/10.1016/j.neubiorev.2019.07.014).
- [95] T Law, BF Malle, and M Scheutz. “A Touching Connection: How Observing Robotic Touch Can Affect Human Trust in a Robot”. English. In: *International Journal of Social Robotics* 13.8 (Dec. 2021), pp. 2003–2019. ISSN: 1875-4791. DOI: [10.1007/s12369-020-00729-7](https://doi.org/10.1007/s12369-020-00729-7).
- [96] J Li, WD Ju, and B Reeves. “Touching a Mechanical Body: Tactile Contact With Body Parts of a Humanoid Robot Is Physiologically Arousing”. English. In: *Journal of Human-Robot Interaction* 6.3 (2017), pp. 118–130. DOI: [10.5898/JHRI.6.3.Li](https://doi.org/10.5898/JHRI.6.3.Li).
- [97] Pei-Chun Lin et al. “An experimental design for facial and color emotion expression of a social robot”. en. In: *The Journal of Supercomputing* 79.2 (Feb. 2023), pp. 1980–2009.

- [98] Yan Liu and Bruno D. Zumbo. “The Impact of Outliers on Cronbach’s Coefficient Alpha Estimate of Reliability: Visual Analogue Scales”. en. In: *Educational and Psychological Measurement* 67.4 (Aug. 2007), pp. 620–634. ISSN: 0013-1644, 1552-3888. DOI: [10.1177/0013164406296976](https://doi.org/10.1177/0013164406296976).
- [99] M. Lorenzini et al. “A synergistic approach to the Real-Time estimation of the feet ground reaction forces and centers of pressure in humans with application to Human-Robot collaboration”. In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 3654–3661. DOI: [10.1109/LRA.2018.2855802](https://doi.org/10.1109/LRA.2018.2855802).
- [100] Marta Lorenzini et al. “Ergonomic Human-Robot Collaboration in Industry: A Review”. In: *Frontiers in Robotics and AI* 9 (2023).
- [101] Dylan P. Losey et al. “A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction”. en. In: *Applied Mechanics Reviews* 70.1 (Jan. 2018), p. 010804. DOI: [10.1115/1.4039145](https://doi.org/10.1115/1.4039145).
- [102] Sylwia Lukasik et al. “The role of socially assistive robots in the care of older people: to assist in cognitive training, to remind or to accompany?” In: *Sustainability* 13.18 (2021), p. 10394.
- [103] Chuhong Luo et al. “Technology Acceptance of Socially Assistive Robots Among Older Adults and the Factors Influencing It: A Meta-Analysis”. In: *Journal of Applied Gerontology* 43.2 (2024), pp. 115–128.
- [104] X. Ma and F. Quek. “Development of a child-oriented social robot for safe and interactive physical interaction”. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS) 2010*. IEEE, 2010, pp. 2163–2168. DOI: [10.1109/IROS.2010.5652715](https://doi.org/10.1109/IROS.2010.5652715).
- [105] A. Madani et al. “Robot-Assisted Drilling on Curved Surfaces with Haptic Guidance under Adaptive Admittance Control”. In: *IEEE International Conference on Intelligent Robots and Systems*. Vol. 2022-October. IEEE, 2022, pp. 3723–3730. DOI: [10.1109/IROS47612.2022.9982000](https://doi.org/10.1109/IROS47612.2022.9982000).
- [106] Federico Manzi et al. “Robots are not all the same: young adults’ expectations, attitudes, and mental attribution to two humanoid social robots”. In: *Cyberpsychology, Behavior, and Social Networking* 24.5 (2021), pp. 307–314.
- [107] U Martinez-Hernandez et al. “Bayesian perception of touch for control of robot emotion”. English. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 4927–4933.

- [108] Zdeněk Materna et al. “Interactive Spatial Augmented Reality in Collaborative Robot Programming: User Experience Evaluation”. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. Aug. 2018, pp. 80–87. DOI: [10.1109/ROMAN.2018.8525662](https://doi.org/10.1109/ROMAN.2018.8525662).
- [109] Iñaki Mautua et al. “Human–robot collaboration in industrial applications: Safety, interaction and trust”. In: *International Journal of Advanced Robotic Systems* 14.4 (July 2017). Publisher: SAGE Publications, p. 1729881417716010. DOI: [10.1177/1729881417716010](https://doi.org/10.1177/1729881417716010).
- [110] Alex Mazursky, Madeleine DeVoe, and Sarah Sebo. “Physical Touch from a Robot Caregiver: Examining Factors that Shape Patient Experience”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 1578–1585. DOI: [10.1109/RO-MAN53752.2022.9900549](https://doi.org/10.1109/RO-MAN53752.2022.9900549).
- [111] Amirhossein H. Memar and Ehsan T. Esfahani. “Objective Assessment of Human Workload in Physical Human-robot Cooperation Using Brain Monitoring”. en. In: *ACM Transactions on Human-Robot Interaction* 9.2 (June 2020), pp. 1–21. DOI: [10.1145/3368854](https://doi.org/10.1145/3368854).
- [112] George Michalos et al. “Seamless human robot collaborative assembly - An automotive case study”. en. In: *Mechatronics* 55 (Nov. 2018), pp. 194–211. DOI: [10.1016/j.mechatronics.2018.08.006](https://doi.org/10.1016/j.mechatronics.2018.08.006).
- [113] R. Midorikawa and M. Niitsuma. “Effects of Touch Experience on Active Human Touch in Human-Robot Interaction”. In: *International Federation of Automatic Control IFAC*. Vol. 51. Issue: 22. Elsevier, 2018, pp. 154–159. DOI: [10.1016/j.ifacol.2018.11.534](https://doi.org/10.1016/j.ifacol.2018.11.534).
- [114] Tracy L Mitzner et al. “Technology adoption by older adults: Findings from the PRISM trial”. In: *The Gerontologist* 59.1 (2019), pp. 34–44.
- [115] S. M. Mizanoor Rahman and Ryojun Ikeura. “Cognition-based variable admittance control for active compliance in flexible manipulation of heavy objects with a power-assist robotic system”. In: *Robotics and Biomimetics* 5.1 (Nov. 2018), p. 7. DOI: [10.1186/s40638-018-0090-x](https://doi.org/10.1186/s40638-018-0090-x).
- [116] Pietro Morasso. “Integrating ankle and hip strategies for the stabilization of upright standing: An intermittent control model”. In: *Frontiers in Computational Neuroscience* 16 (2022).

- [117] Thomas Muller et al. “Transparency Analysis of a Passive Heavy Load Comanipulation Arm”. In: *IEEE Transactions on Human-Machine Systems* (2022). Conference Name: IEEE Transactions on Human-Machine Systems, pp. 1–10. DOI: [10.1109/THMS.2022.3156887](https://doi.org/10.1109/THMS.2022.3156887).
- [118] Keziah Naggita et al. “Parental Responses to Aggressive Child Behavior towards Robots, Smart Speakers, and Tablets”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 337–344. DOI: [10.1109/RO-MAN53752.2022.9900841](https://doi.org/10.1109/RO-MAN53752.2022.9900841).
- [119] Stanislava Naneva et al. “A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots”. en. In: *International Journal of Social Robotics* 12.6 (Dec. 2020), pp. 1179–1201. DOI: [10.1007/s12369-020-00659-4](https://doi.org/10.1007/s12369-020-00659-4).
- [120] JQ Nie et al. “Can You Hold My Hand? Physical Warmth in Human-Robot Interaction”. English. In: *HRI '12: Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2012, pp. 201–202.
- [121] Stefanos Nikolaidis, David Hsu, and Siddhartha Srinivasa. “Human-robot mutual adaptation in collaborative tasks: Models and experiments”. en. In: *The International Journal of Robotics Research* 36.5-7 (June 2017). Publisher: SAGE Publications Ltd STM, pp. 618–634. DOI: [10.1177/0278364917690593](https://doi.org/10.1177/0278364917690593).
- [122] Tatsuya Nomura et al. “Measurement of Anxiety toward Robots”. In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. IEEE, Oct. 2006, pp. 372–377. DOI: [10.1109/ROMAN.2006.314462](https://doi.org/10.1109/ROMAN.2006.314462).
- [123] D Novak et al. “Workload Estimation in Physical Human-Robot Interaction Using Physiological Measurements”. English. In: *Interacting with Computers* 27.6 (Nov. 2015), pp. 616–629. DOI: [10.1093/iwc/iwu021](https://doi.org/10.1093/iwc/iwu021).
- [124] Jordan Nowak et al. “Assistance to Older Adults with Comfortable Robot-to-Human Handovers”. In: *2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*. ISSN: 2162-7576. May 2022, pp. 1–6. DOI: [10.1109/ARSO54254.2022.9802960](https://doi.org/10.1109/ARSO54254.2022.9802960).
- [125] Tian P. S. Oei et al. “Using the Depression Anxiety Stress Scale 21 (DASS-21) across cultures”. In: *International Journal of Psychology* 48.6 (Dec. 2013), pp. 1018–1029.
- [126] M Okuda, Y Takahashi, and S Tsuichihara. “Human Response to Humanoid Robot That Responds to Social Touch”. English. In: *Applied Sciences* 12.18 (Sept. 2022). DOI: [10.3390/app12189193](https://doi.org/10.3390/app12189193).

- [127] PH Orefice et al. “Pressure Variation Study in Human-Human and Human-Robot Handshakes: Impact of the Mood”. English. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2018)*. Ed. by JJ Cabibihan et al. IEEE, 2018, pp. 247–254.
- [128] World Health Organization. *Ageing and health*. en. URL: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (visited on 10/20/2022).
- [129] World Health Organization. *Ageing: Global population*. en. URL: <https://www.who.int/news-room/questions-and-answers/item/population-ageing>.
- [130] Valerio Ortenzi et al. “Object Handovers: A Review for Robotics”. In: *IEEE Transactions on Robotics* 37.6 (Dec. 2021). Conference Name: IEEE Transactions on Robotics, pp. 1855–1873. DOI: [10.1109/TR0.2021.3075365](https://doi.org/10.1109/TR0.2021.3075365).
- [131] Christian Ott et al. “On the passivity-based impedance control of flexible joint robots”. In: *IEEE Transactions on Robotics* 24.2 (2008), pp. 416–429.
- [132] Diego Felipe Paez Granados et al. “Dance Teaching by a Robot: Combining Cognitive and Physical Human–Robot Interaction for Supporting the Skill Learning Process”. In: *IEEE Robotics and Automation Letters* 2.3 (July 2017). Conference Name: IEEE Robotics and Automation Letters, pp. 1452–1459. DOI: [10.1109/LRA.2017.2671428](https://doi.org/10.1109/LRA.2017.2671428).
- [133] Matthew K.X.J. Pan, Elizabeth A. Croft, and Günter Niemeyer. “Evaluating Social Perception of Human-to-Robot Handovers Using the Robot Social Attributes Scale (RoSAS)”. en. In: *2018 ACM/IEEE International Conference on Human-Robot Interaction*. Chicago IL USA: ACM, Feb. 2018, pp. 443–451. DOI: [10.1145/3171221.3171257](https://doi.org/10.1145/3171221.3171257).
- [134] A Pervez and J Ryu. “Safe physical human robot interaction-past, present and future”. English. In: *Journal of Mechanical Science and Technology* 22.3 (Mar. 2008), pp. 469–483. DOI: [10.1007/s12206-007-1109-3](https://doi.org/10.1007/s12206-007-1109-3).
- [135] Luka Peternel et al. “A selective muscle fatigue management approach to ergonomic human-robot co-manipulation”. en. In: *Robotics and Computer-Integrated Manufacturing* 58 (Aug. 2019), pp. 69–79. DOI: [10.1016/j.rcim.2019.01.013](https://doi.org/10.1016/j.rcim.2019.01.013).
- [136] Luka Peternel et al. “Online Human Muscle Force Estimation for Fatigue Management in Human-Robot Co-Manipulation”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. ISSN: 2153-0866. Oct. 2018, pp. 1340–1346. DOI: [10.1109/IROS.2018.8593705](https://doi.org/10.1109/IROS.2018.8593705).

- [137] Luka Peternel et al. “Robot adaptation to human physical fatigue in human–robot co-manipulation”. en. In: *Autonomous Robots* 42.5 (June 2018), pp. 1011–1021. DOI: [10.1007/s10514-017-9678-1](https://doi.org/10.1007/s10514-017-9678-1).
- [138] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. “Guidelines for conducting systematic mapping studies in software engineering: An update”. en. In: *Information and Software Technology* 64 (Aug. 2015), pp. 1–18. DOI: [10.1016/j.infsof.2015.03.007](https://doi.org/10.1016/j.infsof.2015.03.007).
- [139] V Prasad, R Stock-Homburg, and J Peters. “Human-Robot Handshaking: A Review”. English. In: *International Journal of Social Robotics* 14.1 (Jan. 2022), pp. 277–293. DOI: [10.1007/s12369-021-00763-z](https://doi.org/10.1007/s12369-021-00763-z).
- [140] Meiying Qin, Jake Brawer, and Brian Scassellati. “Task-Oriented Robot-to-Human Handovers in Collaborative Tool-Use Tasks”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 1327–1333. DOI: [10.1109/RO-MAN53752.2022.9900599](https://doi.org/10.1109/RO-MAN53752.2022.9900599).
- [141] Akilesh Rajavenkatanarayanan et al. “Towards a robot-based multimodal framework to assess the impact of fatigue on user behavior and performance: a pilot study”. en. In: *12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*. Rhodes Greece: ACM, June 2019, pp. 493–498. DOI: [10.1145/3316782.3322776](https://doi.org/10.1145/3316782.3322776).
- [142] Jacob E. Resch et al. “Balance Performance With a Cognitive Task: A Continuation of the Dual-Task Testing Paradigm”. en. In: *Journal of Athletic Training* 46.2 (Mar. 2011), pp. 170–175. DOI: [10.4085/1062-6050-46.2.170](https://doi.org/10.4085/1062-6050-46.2.170).
- [143] D. Reyes-Uquillas and T. Hsiao. “Safe and intuitive manual guidance of a robot manipulator using adaptive admittance control towards robot agility”. In: *Robotics and Computer-Integrated Manufacturing* 70 (2021). DOI: [10.1016/j.rcim.2021.102127](https://doi.org/10.1016/j.rcim.2021.102127).
- [144] Stephen N. Robinovitch et al. “Effect of Strength and Speed of Torque Development on Balance Recovery With the Ankle Strategy”. en. In: *Journal of Neurophysiology* 88.2 (Aug. 2002), pp. 613–620. DOI: [10.1152/jn.2002.88.2.613](https://doi.org/10.1152/jn.2002.88.2.613).
- [145] Rethink Robotics. *RethinkRobotics*. en. URL: <https://support.rethinkrobotics.com/support/solutions>.
- [146] Matteo Rubagotti et al. “Perceived safety in physical human robot interaction—A survey”. In: *Robotics and Autonomous Systems* (Feb. 2022), p. 104047. DOI: [10.1016/j.robot.2022.104047](https://doi.org/10.1016/j.robot.2022.104047).

- [147] James A Russell and José Miguel Fernández-Dols. *The psychology of facial expression*. en. Cambridge University Press, 1997.
- [148] Melis Sahin and Celal Savur. “Evaluation of Human Perceived Safety during HRC Task using Multiple Data Collection Methods”. In: *2022 17th Annual System of Systems Engineering Conference (SOSE)*. June 2022, pp. 465–470. DOI: [10.1109/SOSE55472.2022.9812693](https://doi.org/10.1109/SOSE55472.2022.9812693).
- [149] Abubakar Tijjani Salihu, Keith D. Hill, and Shapour Jaberzadeh. “Effect of cognitive task complexity on dual task postural stability: a systematic review and meta-analysis”. en. In: *Experimental Brain Research* 240.3 (Mar. 2022), pp. 703–731. DOI: [10.1007/s00221-021-06299-y](https://doi.org/10.1007/s00221-021-06299-y).
- [150] Anara Sandygulova et al. “Individual Differences of Children with Autism in Robot-assisted Autism Therapy”. en. In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Sapporo, Japan: IEEE, Mar. 2022, pp. 43–52. DOI: [10.1109/HRI53351.2022.9889537](https://doi.org/10.1109/HRI53351.2022.9889537).
- [151] Juergen Sauer, Andreas Sonderegger, and Sven Schmutz. “Usability, user experience and accessibility: towards an integrative model”. In: *Ergonomics* 63.10 (Oct. 2020), pp. 1207–1220. DOI: [10.1080/00140139.2020.1774080](https://doi.org/10.1080/00140139.2020.1774080).
- [152] J. Schmidler et al. “A questionnaire for the evaluation of physical assistive devices (quead): Testing usability and acceptance in physical human-robot interaction”. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*. Vol. 2017-January. IEEE, 2017, pp. 876–881. DOI: [10.1109/SMC.2017.8122720](https://doi.org/10.1109/SMC.2017.8122720).
- [153] Schönbrodt, Perugini Felix D., and Marco. “At what sample size do correlations stabilize?” In: *Journal of Research in Personality* 47 (2013), pp. 609–612.
- [154] Roxana Schweighart et al. “Wishes and Needs of Nursing Home Residents”. en. In: *Healthcare* 10.5 (May 2022). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 854.
- [155] M. Selvaggio et al. “Autonomy in physical human-robot interaction: A brief survey”. In: *IEEE Robotics and Automation Letters* 6.4 (2021), pp. 7989–7996. DOI: [10.1109/LRA.2021.3100603](https://doi.org/10.1109/LRA.2021.3100603).
- [156] M Shiomi and N Hagita. “Audio-Visual Stimuli Change not Only Robot’s Hug Impressions but Also Its Stress-Buffering Effects”. English. In: *International Journal of Social Robotics* 13.3 (June 2021), pp. 469–476. DOI: [10.1007/s12369-019-00530-1](https://doi.org/10.1007/s12369-019-00530-1).

- [157] M Shiomi and N Hagita. “Do Audio-Visual Stimuli Change Hug Impressions?” English. In: *Social Robotics, ICSR 2017*. Ed. by A Kheddar et al. Vol. 10652. Springer Nature, 2017, pp. 345–354. DOI: [10.1007/978-3-319-70022-9_34](https://doi.org/10.1007/978-3-319-70022-9_34).
- [158] M Shiomi, T Minato, and H Ishiguro. “Effects of Robot’s Awareness and its Subtle Reactions Toward People’s Perceived Feelings in Touch Interaction”. English. In: *Journal of Robotics and Mechatronics* 32.1 (Feb. 2020), pp. 43–50.
- [159] M Shiomi et al. “Gaze-Height and Speech-Timing Effects on Feeling Robot-Initiated Touches”. English. In: *Journal of Robotics and Mechatronics* 32.1 (Feb. 2020), pp. 68–75.
- [160] M Shiomi et al. “How Should a Robot React Before People’s Touch?: Modeling a Pre-Touch Reaction Distance for a Robot’s Face”. English. In: *IEEE Robotics and Automation Letters* 3.4 (Oct. 2018), pp. 3773–3780. DOI: [10.1109/LRA.2018.2856303](https://doi.org/10.1109/LRA.2018.2856303).
- [161] Linda Shore, Adam de Eyto, and Leonard O’Sullivan. “Technology acceptance and perceptions of robotic assistive devices by older adults—implications for exoskeleton design”. In: *Disability and rehabilitation: assistive technology* 17.7 (2022), pp. 782–790.
- [162] Bruno Siciliano et al. *Robotics*. en. Springer London, 2009.
- [163] Vésteinn Sigurjónsson, Kerstin Johansen, and Carin Rösiö. “Exploring the operator’s perspective within changeable and automated manufacturing – A literature review”. en. In: *Procedia CIRP* 107 (2022), pp. 369–374.
- [164] Ana Correia Simões et al. “Designing human-robot collaboration (HRC) workspaces in industrial settings: A systematic literature review”. en. In: *Journal of Manufacturing Systems* 62 (Jan. 2022), pp. 28–43. DOI: [10.1016/j.jmsy.2021.11.007](https://doi.org/10.1016/j.jmsy.2021.11.007).
- [165] Valentina De Simone et al. “Human-Robot Collaboration: an analysis of worker’s performance”. en. In: *Procedia Computer Science* 200 (2022), pp. 1540–1549. DOI: [10.1016/j.procs.2022.01.355](https://doi.org/10.1016/j.procs.2022.01.355).
- [166] Human Factors & Ergonomics Society. *What is Human Factors and Ergonomics — HFES*. URL: <https://www.hfes.org/About-HFES/What-is-Human-Factors-and-Ergonomics>.
- [167] Kyle Wayne Strabala et al. “Towards Seamless Human-Robot Handovers”. en. In: *Journal of Human-Robot Interaction* 2.1 (Mar. 2013), pp. 112–132. DOI: [10.5898/JHRI.2.1.Strabala](https://doi.org/10.5898/JHRI.2.1.Strabala).

- [168] Jefferey Too Chuan Tan et al. “Safety Strategy for Human–Robot Collaboration: Design and Development in Cellular Manufacturing”. en. In: *Advanced Robotics* 24.5-6 (Jan. 2010), pp. 839–860. DOI: [10.1163/016918610X493633](https://doi.org/10.1163/016918610X493633).
- [169] Mohsen Tavakol and Reg Dennick. “Making sense of Cronbach’s alpha”. In: *International Journal of Medical Education* 2 (June 2011), pp. 53–55. ISSN: 2042-6372. DOI: [10.5116/ijme.4dfb.8dfd](https://doi.org/10.5116/ijme.4dfb.8dfd).
- [170] G. Trovato et al. “Is hugging a robot weird? Investigating the influence of robot appearance on users’ perception of hugging”. In: *IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2016, pp. 318–323. DOI: [10.1109/HUMANOIDS.2016.7803295](https://doi.org/10.1109/HUMANOIDS.2016.7803295).
- [171] M. Y. Tsalamlal et al. “Affective handshake with a humanoid robot”. In: *2015 ACII*. Sept. 2015, pp. 334–340.
- [172] Emily Tweel, Arnold J. Stromberg, and Geetanjali Gera. “Use of Wearable Sensors to Assess the Effects of Performing a Cognitive Task on Sensory Integration of Balance in Healthy Individuals”. en. In: *Sensors* 22.7 (Jan. 2022). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 2776. DOI: [10.3390/s22072776](https://doi.org/10.3390/s22072776).
- [173] Jacqueline Urakami and Katie Seaborn. “Nonverbal Cues in Human-robot Interaction”. In: *ACM Trans* (Dec. 2022).
- [174] Christiaan H. Vinkers et al. “The effect of stress on core and peripheral body temperature in humans”. In: *Stress* 16.5 (Sept. 2013), pp. 520–530. DOI: [10.3109/10253890.2013.807243](https://doi.org/10.3109/10253890.2013.807243).
- [175] Qiao Wang et al. “Computational Model of Robot Trust in Human Co-Worker for Physical Human-Robot Collaboration”. In: *IEEE Robotics and Automation Letters* 7.2 (Apr. 2022). Conference Name: IEEE Robotics and Automation Letters, pp. 3146–3153. DOI: [10.1109/LRA.2022.3145957](https://doi.org/10.1109/LRA.2022.3145957).
- [176] Xueqin Wang et al. “An investigation of self-service technology usage during the COVID-19 pandemic: The changing perceptions of ‘self’ and technologies”. In: *Technology in society* 70 (2022), p. 102032.
- [177] Yiwei Wang et al. “Quantitative Modeling and Analysis of Reliance in Physical Human–Machine Coordination”. In: *Journal of Mechanisms and Robotics* 11.6 (Sept. 2019). DOI: [10.1115/1.4044545](https://doi.org/10.1115/1.4044545).
- [178] CJAM Willemse and JBF van Erp. “Social Touch in Human-Robot Interaction: Robot-Initiated Touches can Induce Positive Responses without Extensive Prior Bonding”. English. In: *International Journal of Social Robotics* 11.2 (Apr. 2019), pp. 285–304. DOI: [10.1007/s12369-018-0500-9](https://doi.org/10.1007/s12369-018-0500-9).

- [179] CJAM Willemse et al. “Observing Touch from Video: The Influence of Social Cues on Pleasantness Perceptions”. English. In: *Haptics: Perception, Devices, Control, and Applications, EuroHaptics 2016, Part II*. Ed. by F Bello, H Kajimoto, and Y Visell. Vol. 9775. Springer International Publishing, 2016, pp. 196–205. DOI: [10.1007/978-3-319-42324-1_20](https://doi.org/10.1007/978-3-319-42324-1_20).
- [180] Claes Wohlin. “Guidelines for snowballing in systematic literature studies and a replication in software engineering”. en. In: *2014 18th International Conference on Evaluation and Assessment in Software Engineering - (EASE)*. London, England, United Kingdom: ACM Press, 2014, pp. 1–10. DOI: [10.1145/2601248.2601268](https://doi.org/10.1145/2601248.2601268).
- [181] Y. Yamada, Y. Umetani, and Y. Hirasawa. “Proposal of a psychophysiological experiment system applying the reaction of human pupillary dilation to frightening robot motions”. In: *IEEE SMC’99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics*. Vol. 2. ISSN: 1062-922X. Oct. 1999, 1052–1057 vol.2. DOI: [10.1109/ICSMC.1999.825408](https://doi.org/10.1109/ICSMC.1999.825408).
- [182] Y Yamashita et al. “Path Analysis for the Halo Effect of Touch Sensations of Robots on Their Personality Impressions”. English. In: *Social Robotics, ICSR 2016*. Ed. by A Agah et al. Vol. 9979. Springer Nature, 2016, pp. 502–512. DOI: [10.1007/978-3-319-47437-3_49](https://doi.org/10.1007/978-3-319-47437-3_49).
- [183] A Yazdani et al. “DULA and DEBA: Differentiable Ergonomic Risk Models for Postural Assessment and Optimization in Ergonomically Intelligent pHRI”. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9124–9131. DOI: [10.1109/IROS47612.2022.9981528](https://doi.org/10.1109/IROS47612.2022.9981528).
- [184] N Zamani et al. “Effects of Motion Parameters on Acceptability of Human-Robot Patting Touch”. English. In: *2020 IEEE Haptics Symposium (HAPTICS)*. IEEE, 2020, pp. 664–670. DOI: [10.1109/HAPTICS45997.2020.ras.HAP20.36.d8bb0c58](https://doi.org/10.1109/HAPTICS45997.2020.ras.HAP20.36.d8bb0c58).
- [185] XQ Zheng et al. “How Can Robots Make People Feel Intimacy Through Touch?” English. In: *Journal of Robotics and Mechatronics* 32.1 (Feb. 2020), pp. 51–58.
- [186] XQ Zheng et al. “Modeling the Timing and Duration of Grip Behavior to Express Emotions for a Social Robot”. English. In: *IEEE Robotics and Automation Letters* 6.1 (Jan. 2021), pp. 159–166. DOI: [10.1109/LRA.2020.3036372](https://doi.org/10.1109/LRA.2020.3036372).
- [187] XQ Zheng et al. “What Kinds of Robot’s Touch Will Match Expressed Emotions?” English. In: *IEEE Robotics and Automation Letters* 5.1 (Jan. 2020), pp. 127–134. DOI: [10.1109/LRA.2019.2947010](https://doi.org/10.1109/LRA.2019.2947010).

- [188] Mark Zolotas et al. “Productive Inconvenience: Facilitating Posture Variability by Stimulating Robot-to-Human Handovers”. en. In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. Napoli, Italy: IEEE, Aug. 2022, pp. 122–128. DOI: [10.1109/RO-MAN53752.2022.9900666](https://doi.org/10.1109/RO-MAN53752.2022.9900666).

APPENDICES

Appendix A

Questionnaires

A.1 StuMMB-Q

Below is the adapted version of the [StuMMB-Q](#) questionnaire.

We would like to learn how difficult you found the completed math problems while simultaneously maintaining your balance on the board. When responding to the following questions, please consider the overall experience of solving the math problems while balancing on the board, rather than focusing solely on the math problems.

On a scale from 1 (Not at all) to 3 (Totally), rate how strongly the following statements are true for you:

1. †The tasks were difficult to answer.
2. * At the processing of the tasks I haven't done my best particularly.
3. †The contents of the tasks were complicated.
4. †The tasks were challenging.
5. * I haven't taken particular trouble with the reply to the tasks.
6. †* The tasks were easy to work on.
7. I have made an effort at the processing of the tasks.
8. †* The contents of the tasks were easy to understand.

9. At the reply to the tasks I have made an effort intellectually.
10. †* The tasks were easy to solve.
11. * I haven't particularly focused when solving the tasks.
12. I have given my best to complete the tasks.

Items with a dagger (†) contribute to Mental Load score, while item without a dagger contribute to Mental Effort score. Items marked with an asterisk (*) are coded in reverse.