Robot Trust and Self-Confidence Role Arbitration Method for Physical Human-Robot Collaboration

by

QIAO WANG

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisor: Dr. Marc Carmichael

Co-supervisor: Dist. Prof. Dikai Liu

Co-supervisor: Dist. Prof. Chin-Teng Lin

at the

Robotics Institute

Faculty of Engineering and Information Technology

University of Technology Sydney

18th April 2024

Certificate of Original Authorship

I, Qiao Wang declare that this thesis, is submitted in fulfilment of the requirements for the award of a Doctor of Philosophy degree, in the School of Mechanical and Mechatronic Engineering at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signed

30/04/2024

Date

Abstract

The emergence of collaborative robots (cobots) allows humans and robots to work in the same space. By collaborating on a common task, the strengths of humans (e.g. perception and decision-making) and robots (e.g. repeatability and accuracy) can be combined to achieve better performance. Human co-workers could interact with a robot in various ways and in various applications such as rehabilitation and materials handling. This thesis focusses on physical Human-Robot Collaboration (pHRC). pHRC can be defined as a human co-worker physically contacting or exchanging force continuously with a robot to accomplish a shared goal in the same workspace.

Trust is the key to achieving successful Human-Robot Interaction (HRI). Trust in robotics is defined by the degree to which humans are willing to utilize robots. Factors such as a robot's functionality, reliability, and features, along with the type of task and the environment, can affect trust levels. Inappropriate trust levels can result in excessive or insufficient reliance on robots, due to overly high or low trust, respectively, leading to poor teamwork. As a result, this can increase both physical and cognitive workloads and decrease overall task performance, including accuracy and completion time. In addition to the human co-worker trusting the robot, the trust of the robot in its human co-worker should also be considered.

In this Thesis, a computational model of a robot's trust in its human co-worker for pHRC is proposed. The trust model is a function of the human co-worker's performance, which can be characterised by factors including safety, robot singularity, smoothness, physical performance, and cognitive performance. Safety performance is defined based on the possibility of collision between the robot and surrounding objects. Singularity performance is proposed to quantify human performance in avoiding singular configurations. Smoothness performance is used to quantify the degree of smoothness when the human co-worker moves the robot during performing a pHRC task. Physical performance is based on an estimate of the human co-worker to physically contribute to the task. Cognitive performance is based on an estimate of cognitive capacity while performing a task. The proposed computational trust model provides a comprehensive evaluation of human co-worker performance

in pHRC. The model could be used to monitor human performance in real-time and adapt the behaviour of a system to improve human-robot combined collaboration efficiency and experience.

Moreover, role arbitration in human-robot collaboration (HRC) is a dynamically changing process that is affected by many factors, such as physical workload, environmental changes, and trust. To address this dynamic process, a trust-based role arbitration method is studied in this thesis. A computational model of robot trust and self-confidence (TSC) in pHRC is proposed. The proposed TSC model is defined as the difference between robot-to-human trust and the robot's self-confidence. The robot-to-human trust model is established as previously mentioned. Then, a role arbitration method is proposed based on the presented TSC model. The robot's self-confidence is modelled based on whether a human agrees with the control actions of the robot. The robot's self-confidence is high when a human agrees with robot control and vice versa. In addition, when the robot's trust in a human co-worker is much higher than the robot's self-confidence, the control is allocated to the human co-worker and vice versa. However, the way the control is allocated when the robot's trust in human co-workers and itself are both high or low becomes a question. In this work, it is proposed that the history of robot self-confidence can be used. If the robot's self-confidence in history is high, it indicates that the robot control was satisfied by the human in the past. Hence, the control will be biased toward the robot and vice versa. Human-in-the-loop experiments with a collaborative robot are conducted to verify the TSC-based role arbitration method. The results show that the proposed method could achieve superior human-robot combined performance, reduce human co-workers' workload, and improve subjective preference.

Furthermore, the Rating Scale (RS) method has traditionally been the go-to approach for assessing subjective experiences. Yet, its suitability for physical human-robot collaboration (pHRC) research is questionable due to several inherent issues, such as response bias, individual differences among participants, and the challenges posed by the scale's level of detail. Variabilities between individuals can significantly skew the results obtained by rating scales. Scales with too many options might confuse participants, causing ambiguous and potentially biased answers, whereas scales with too few options risk overlooking the subtle complexities of human emotions. Moreover, getting indifferent replies can risk the trustworthiness of the data collected.

In light of these obstacles, this thesis advocates the use of Pairwise Comparison (PC) in pHRC research, which is a comparative survey method that pits items against each other based on specific criteria. Inspired by the structure of the NASA Task Load Index (NASA-TLX), both RS and PC surveys are crafted and employed in a sequence of pHRC experiments. The results indicate that PC outperforms the rating scale in terms

of precision and stability. By offering straightforward questions and potentially shortening the duration of experiments, PC not only maintains participant engagement but also ensures consistent accuracy and reliability across varying experimental designs. More robust measurement of interaction experience in pHRC could improve the design of pHRC systems.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisory panel, Dr. Marc Carmichael, Prof. Dikai Liu, and Prof Chin-Teng Lin, for their untiring support, guidance, and mentorship throughout my human-robot interaction Ph.D. journey. Their invaluable expertise and insights have shaped the direction and quality of my research.

Second, I would also like to extend my appreciation to my fellow lab members and research colleagues for their camaraderie, inspired discussions, and willingness to share ideas. A special thanks to Ziqi Wang, Jonathan Woolfrey, Sheila Sujito, Yujun Lai, Yingyu Wang, Tiancheng Li, and Yang Song for their unwavering friendship and support during challenging times.

My heartfelt thanks go to the administrative and technical staff at Robotics Institute for their assistance and for providing a conducive environment for my research.

I am grateful for the financial support provided by the Australian Research Council (ARC) Discovery Project Grant [DP210101093], which enabled me to pursue my doctoral studies.

Lastly, I would like to express my profound appreciation to my family, particularly my parents, Linzhong Wang, and Jinyu Zhai, for their unconditional love, support, and belief in me. Without their encouragement, I would not have been able to reach this milestone.

Contents

List of Figures

List of Tables

Abbreviations

RI Robotics Institute

UTS Univeristy of Technology Sydney

HHI Human-Human Interaction

HRI Human-Robot Interaction

HRC Human-Robot Collaboration

pHRC Physical Human-Robot Collaboration

WMSDs Work-related musculoskeletal disorders

EMG Electromyography

EEG Electroencephalogram

T Robot Trust in Human Co-worker

SC Robot Self-Confidence

TSC Robot Trust and Self-Confidence

RA Robot Trust and Self-confidence based role arbitration method

HL Human-Leading: Human in control

RL Robot-Leading: Robot in control

PC Pairwise comparison subjective evaluation method

RS Rating scale subjective evaluation Method

RMT Designed red moving target for the experiment design

MD Mental Demand in NASA-TLX

PD Physical Demand in NASA-TLX

TD Temporal Demand in NASA-TLX

EF Effort in NASA-TLX

PER Performance in NASA-TLX

FRU Frustration in NASA-TLX

Nomenclature

T	robot's trust in human co-workers.
p_H	Human co-worker performance.
p_R	Robot performance.
β	a discount factor is introduced that represents the effect of the past performance.
p_A^c	Critical performance factor.
p_A^{nc}	Non-critical performance factor.
C	The maximum contribution of non-critical performance factor.
γ_A	weighting ratio of non-critical performance factors.
p_S	Safety Performance.
a^-	Lower threshold of safety.
a^+	Upper threshold of safety.
p_{SP}	Singularity Performance.
σ_{AW}^+	Upper threshold of the smallest singular value when away from the singular configuration.
σ_{AW}^-	Lower threshold of the smallest singular value when away from the singular configuration.
σ_{AP}^+	Upper threshold of the smallest singular value when approaching the singular configuration.
σ_{AP}^-	Lower threshold of the smallest singular value when approaching the singular
p_{SM}	Smoothness Performance.
sm^-	Lower threshold of smoothness.
sm^+	Upper threshold of smoothness.
p_{PW}	Physical Performance.
$F_{max,iso}$	Maximum isometric force.
MVC	Maximum voluntary contraction.
C_f	Fatigue coefficient.
C_r	Recovery coefficient.
p_{CP}	Cognitive Performance.

p_{CP}^{+}	Upper threshold of Cognitive Performance.
p_{CP}^{-}	Lower threshold of Cognitive Performance.
β_{CP}	Task difficulty.
r	Utilization ratio.
a_r	sensitivity of the next time step utilization ratio
$\dot{x}$	the velocity command sent to the robot manipulator
$\dot{x}_H$	Human control
$\dot{x}_R$	Robot control
M_d	Desired robot inertia matrix
M_d	Desired robot damping matrix
F_H	Magnitude of human force interact with robot measured through force torque sensor
α	Role arbitration function
V_D	Difference between human and robot control
TSC^{+}	Upper threshold of robot trust and self-confidence TSC
TSC^{-}	Lower threshold of robot trust and self-confidence TSC

Chapter 1

Introduction

1.1 Background

The ever-evolving landscape of technology, particularly in areas such as sensors and actuators, has paved the way for a new era in human-robot interaction. This shift is marked by the emergence of collaborative robots, or "cobots", which are explicitly designed to co-exist with humans in a shared workspace, as opposed to their conventional industrial robot counterparts operating in isolation [9].

The collaboration between humans and robots aims to harness the unique strengths of each party. Although robots often possess superior power, endurance, and payload capabilities, humans exhibit exceptional cognitive abilities, including adaptability, dexterity, and aptitude in handling complex and uncertain situations. This collaborative approach is particularly beneficial in real-world environments characterised by unpredictability and dynamism, where robots cannot perform tasks independently.

One of the primary goals of human-robot collaboration is to alleviate the burden on human co-workers by reducing the physical and cognitive demands placed on them. In addition, it strives to enhance the combined performance of human-robot teams, potentially exceeding the capabilities of either party alone.

There are various ways in which humans could actively interact with cobots. Physical human-robot interaction (pHRI) is one of the interactions. This thesis focusses on the physical Human-Robot Collaboration (pHRC), a subset of pHRI. Collaboration can be defined as "working jointly with others or together, especially in an intellectual endeavour" [10]. Both parties in the human-robot dyad share knowledge of the goal of the pHRC task. Therefore, pHRC defines a human co-worker as physical contacts or exchanging force

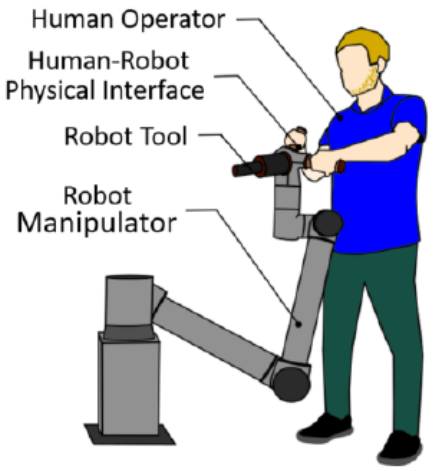


FIGURE 1.1: A pHRC scenario, with a human co-worker physically controlling a robot manipulator [1]

with a robot to accomplish a shared goal in the same workspace. The robot motions are wholly or partially under control from the human co-worker through continuous force exchange. An example of the pHRC scenario examined in this thesis is shown in Figure 1.1, in which a human co-worker operates the robot manipulator by physically moving the handlebar mounted on the robot end effector. Admittance and Impedance control are control paradigms commonly used in pHRC where the relationship between the force exerted by the robot and the velocity of its end-effector is defined. For example, admittance control uses external forces and torques as input signals and produces desired velocities as output. This approach can achieve a safe and compliant interaction between the robot and its environment, making it suitable for tasks that require physical contact or collaboration with humans.

A human-robot dyad can use pHRC to help human co-workers perform demanding tasks by leveraging the power and accuracy advantages of the robot, whilst maintaining the benefits of the human co-worker's superior cognitive, perception and decision-making abilities. These characteristics can bring benefits to various applications, such as surgical robots [11], rehabilitation [12–14], and material handling [1, 15].

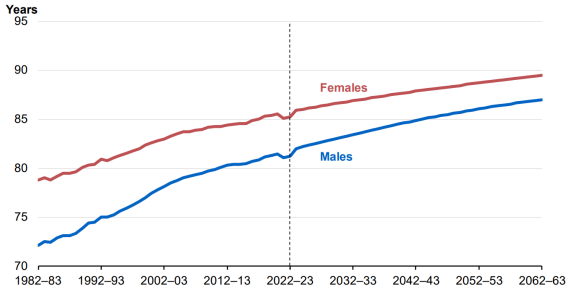
1.2 Challenge and Motivation

1.2.1 Aging issue

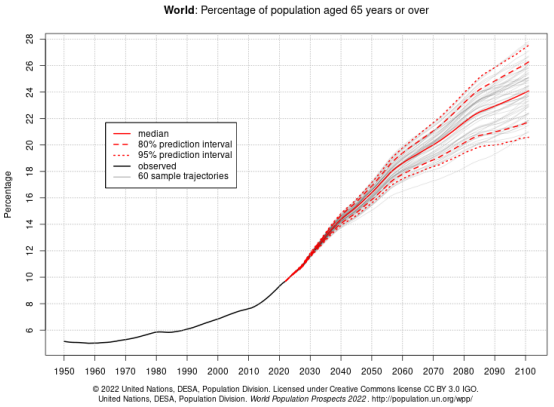
Due to the constant improvement in technology and social welfare, the life expectancy as shown in Figure 1.2a [16] of the population continues to increase. In the next 40 years,

the number of people 65 years old will double and the number of 85 years old will triple in Australia. According to the United Nations demographic statistical report, the percentage of the population over 65 years of age is shown in Figure 1.2b [17]. From the figure, the percentage of the population over 65 years old will increase from 10% to 16% (2050) to a quarter of the population at the end of the century. It is clear an ageing population will become an unavoidable and urgent issue that needs to be considered. A shrinking labour force will become a serious problem that affects the economy and living standards.

Therefore, the development of robotic technology is essential in the future. pHRC can help mitigate this issue by increasing worker productivity, allowing the elderly population to remain in the workforce for longer, and improving healthcare.



(a) Life expectancy at birth for females and males in Australia [16]



(b) Percentage of population aged 65 or over in worldwide [17]

FIGURE 1.2: Statistical Evidence for aging issue worldwide and Australia: (a) Percentage of population aged over 65 worldwide and (b) life expectancy at birth for females and males in Australia.

1.2.2 Industrial Occupational Health

Work-related musculoskeletal disorders (WMSDs) have become a prevalent problem for industrial workers due to the long-term impact of heavy workload and repetitive tasks. The construction industry is synonymous with occupational risks, particularly musculoskeletal disorders (MSDs), which are an important problem that leads to loss of productivity, functional impairments, and even permanent disability [18]. The risk of MSDs varies between construction professions due to distinct biomechanical factors. Although primary prevention efforts have reduced the biomechanical load in many roles, long-term studies reveal that MSDs continue to be prevalent. A recent study found that more than half of active construction workers experienced musculoskeletal complaints, which significantly impacts

their career longevity, as shown in Figure 1.3a. In addition to the loss of productivity from industry and construction workers, on average the loss of time a day is around 1.8 hours per day to 2.2 hours per day, as shown in Figure 1.3b. MSD could also result in a significant reduction in life quality, increased disability, and reduced workers' performance [19–21]. This will result in a high medical burden on individuals or the country [22].

	Prevalence (%)	Median (IQR[a])
Demographic factors		
Age		41 (30–51)
Having a partner		
No	22	
Yes, without a paid job	24	
Yes, with a paid job	54	
Work-related factors		
Shift work	12	
Overtime work	66	
Dangerous work	63	
Physical job demands		
Using force		
No	16	
Occasionally	37	
Frequently	47	
Working in awkward postures		
No	21	
Occasionally	46	
Frequently	33	
Exposure to vibrations		
No	29	
Occasionally	33	
Frequently	38	
Psychosocial factors		
Quantitative job demands (1–4)		2.3 (1.8–2.5)
Job autonomy (1–3)		2.6 (2.2–3.0)
Skill discretion (1–4)		3.0 (2.3–3.3)
Emotional job demands (1–4)		1.3 (1.0–2.0)
Co-worker support (1–4)		3.0 (3.0–3.5)
Social support supervisor (1–4)		3.0 (2.5–3.0)
Health-related factors		
Emotional exhaustion	13	
Musculoskeletal symptoms		
Never	42	
Occasional	37	
Frequent	21	
Ability and willingness to continue working until the age of 65		
Ability	30	
Willingness	29	

[a] Interquartile range (25th–75th percentile)

(a) Survey of the construction worker (n = 5610) under work-related musculoskeletal disorders

	Industrial workers (n = 388)	Construction workers (n = 182)
Health and Labor Questionnaire [13]		
Limitations due to health problems	0.217	0.17
Limitations due to MSD problems	0.097	0.113
Lost work time due to health problems	0.053	0.065
Lost work time due to MSD problems	0.012	0.03
Average hours/day lost due to health problems[a]	1.0 hr/day	1.5 hr/day
Average hours/day lost due to MSD problems[a]	0.5 hr/day	2.0 hr/day
Quality and Quantity Questionnaire [8]		
Health problems during previous work day	0.205	0.244
MSD problems during previous work day	0.109	0.209
Lost work time due to health problems	0.107	0.118
Lost work time due to MSD problems	0.045	0.089
Average hours/day lost due to health problems[a]	2.2 hr/day	1.8 hr/day
Average hours/day lost due to MSD problems[a]	1.9 hr/day	1.9 hr/day

Abbreviations: MSD, musculoskeletal disorders.
[a] Among those with lost work time.

(b) Productivity loss at work due to health problem among industrial workers and construction workers.

FIGURE 1.3: Statistical Evidence for Work-related musculoskeletal disorders (WMSDs) in Industry: (a) Survey of the construction worker (n = 5610) under work-related musculoskeletal disorders, including 316 painters, 1030 plumbers, welders and fitters, 1072 electricians and assemblers, and other construction worker [23]. (b) Productivity loss at work due to health problem among industrial workers and construction workers [24].

To address this issue, pHRC is critical due to the strength and combined accuracy advantage of the robot.

1.2.3 pHRC in Industrial Applications

A team from the University of Technology Sydney developed a new robotic system ANBOT (Assistance as Needed roBOT) that addresses occupational health problems in the abrasive blasting industry [1]. In conventional industrial blasting such as that shown in Figure 1.4a, loads as large as 100N are constantly applied to the human co-worker, resulting in non-ergonomic situations during practice. Furthermore, strong vibrations are another major factor that causes WMSD in workers during abrasive blasting operations. The ANBOT could handle most of the workload brought by the reaction forces imposed on the worker during blasting. The worker only requires a small amount of effort to manipulate the nozzle, as shown in Figure 1.4b, which has been validated in field tests for the worker who has no experience working with the robot.

(a) Manual Abrasive Blasting

(b) ANBOT with Human Co-worker on Site Trial

FIGURE 1.4: (a) An typical manual industrial abrasive blasting task [1] (b) ANBOT with a human co-worker in the site trial [2]

1.2.4 pHRC in Medicare Applications

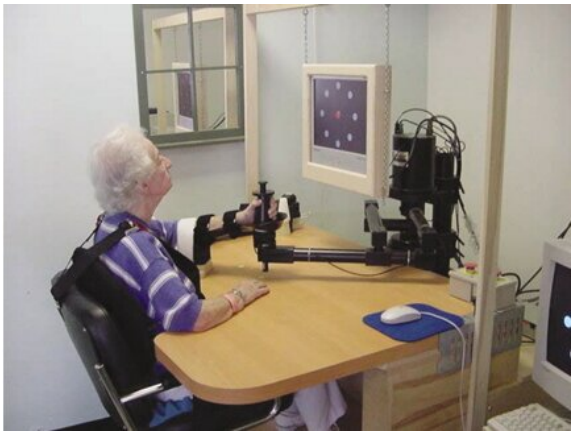
Assistive Robotics are intelligent machines designed to work collaboratively with humans in roles ranging from assistants to companions. A prime example of such innovation is the 'Smart Hoist as shown in Figure 1.5, a patient lifting device equipped with sensors and powered wheels [25]. While it functions as an advanced patient lifter, it significantly eases operational effort, potentially reducing lower back injuries among caregivers. Preliminary

evaluations of the IRT Woonona residential care facility suggest its ease of use and potential to improve caregiver safety.

FIGURE 1.5: 'Smart Hoist': a patient lifting device equipped with sensors and powered wheels. Although functioning as a advanced patient lifter, it significantly eases the operational effort, potentially reducing lower back injuries among caregivers.

The growing scarcity of therapists and caregivers to help people with physical disabilities at home is expected to intensify, posing significant challenges in the near future due to an ageing problem. There is also an increasing number of patients who require physical rehabilitation of the upper extremities. Although recent studies highlight the promise of robotic devices in addressing these concerns, their presence in clinical settings remains minimal, indicating substantial opportunities for improvement. A study in [3] presents a novel approach to robotic therapy, termed performance-based progressive robot therapy, which initiates robot assistance according to specific criteria such as speed, time, or EMG readings as shown in Figure 1.6a. Given the high incidence of stroke-related disabilities in the US, the research focusses primarily on this group. Research, which includes extensive clinical trials with more than 200 stroke patients, underscores the effectiveness of repetitive robotic-assisted techniques in alleviating post-stroke arm problems. The study is also keen to tailor therapy to individual needs, with significant exploration of an impedance control algorithm driven by these specific criteria. Initial feedback from therapists has indicated a considerable benefit of this method: a significant reduction in arm rigidity.

Lower limb rehabilitation, illustrated in Figure 1.6b, represents a prominent area of research that has garnered significant attention among scholars. The primary goal in this field is to reinstate walking abilities in individuals affected by brain and spinal cord injuries. A groundbreaking development in this arena is the electromechanical gait tool, LokoHelp, engineered by Medburg Basel. This innovative device has been evaluated for its effectiveness in training outcomes, as well as the comfort and exertion levels experienced by both patients and therapists, as discussed in [4].



(a) Stroke patient during robot-aided therapy

(b) Spinal cord injuries for the purpose of restoring the walking capability

FIGURE 1.6: Examples of pHRC in Medicare applications in rehabilitation robotics: (a) Stroke patient during robot-aided therapy [3]. (b) LokoHelp: A rehabilitation robot for brain and spinal cord injuries for the purpose of restoring walking capability [4].

1.3 Trust in HRC

Social Human-Robot Interaction (sHRI) encompasses the interdisciplinary study, design, and assessment of robots designed to interact with humans in ways that are socially significant and meaningful. Within the realm of sHRI, a multitude of cognitive factors critically shapes the nature and success of human-robot engagements. These include, but are not limited to: situation awareness, attention and focus, decision-making processes, reliability, and trust. These elements not only enrich the interaction dynamics but also significantly determine the effectiveness and outcomes of these social exchanges [26].

Trust is a pivotal component of the human cognitive process. As defined in [27], trust can be described as the belief that an entity, be it a machine or another individual, will assist in accomplishing one's goals, especially in contexts filled with ambiguity and risk. Inappropriate levels of trust in a robot can lead to either disuse (under-reliance) or misuse (over-reliance) as indicated in [28]. In human-human interactions (HHI), trust operates bidirectionally. To mirror this in human-robot interactions (HRI), it is crucial to embed cognitive functions within robots that not only include trust in humans but also embody a form of robotic self-assurance. This implies that robots should extend beyond the basic sense-plan-act paradigm and be instilled with human-like cognitive characteristics, like a semblance of 'self-awareness', akin to the aforementioned trust and self-confidence as shown in Figure 1.7.

Trust in HRI has been the subject of thorough investigation. Subjective questionnaires are a widely used method to assess trust from humans towards robots, provided by researchers during experiments as noted in the work by Muir [29]. However, this approach allows for the collection of questionnaire outcomes only post-experiment or necessitates pausing the experiment for data collection, making it unsuitable for real-time computational modeling and robot control. This limitation underscores the need for developing real-time computational trust models that enable robots to adjust their behavior dynamically based on trust levels, aiming to enhance the joint performance of humans and robots and to alleviate the physical and cognitive load on human coworkers.

The pioneering study on computational models of human trust in robots was conducted by Lee and Moray [5], identifying faults and performance as key factors influencing trust variability. They crafted a dynamic trust model using an autoregressive moving average model that incorporates operators' trust ratings, along with the performance and faults of the automatic controller. Xu [6] introduced an online probabilistic trust model that uses a dynamic Bayesian network to calculate trust based on the history of interactions.

Trust is a two-way street, involving both the human and the robot coworkers during interactions. Research has also been done on models of robot trust in humans. For instance, Rahman [7] adapted the human-to-robot trust model by Lee and Moray [5] for tasks involving handovers, allowing robots to adjust their handover strategy based on their trust in humans. Similarly, Tran [8] developed a model of robot-to-human confidence for tasks like grit-blasting in pHRC, utilizing a fluid-stochastic Petri net model.

While human-to-robot trust models [5, 6] have relied on subjective questionnaires to define their parameters, such a method is not viable for creating robot-to-human trust models, given that robots lack an inherent sense of trust to be surveyed. In [7] and [8], these models are informed by performance metrics, like the blasting angle or the variability in blasting paths, though their application has been task-specific which is not comprehensive.

1.4 Role Arbitration in HRC

In many pHRC applications, the human primarily assumes the leadership role within the system. However, autonomous systems excel in structured settings, such as logistics warehouses or automotive manufacturing, as shown in Figure 1.8. Here, they can efficiently execute repetitive, high-volume tasks with minimal errors and impressive precision, thanks to their advanced sense-plan-act mathematical model. However, autonomous systems face challenges, including varying lighting conditions and occlusions. These challenges are exacerbated in human-robot interaction scenarios, such as hospitals or industrial situations,

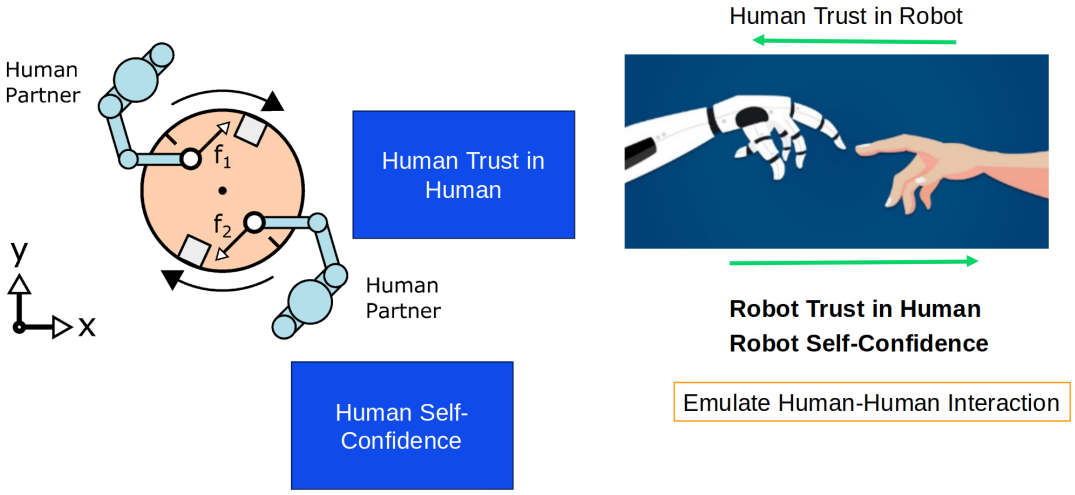


FIGURE 1.7: Emulate Human-Human Interaction paradigm into human-robot interaction

where human involvement introduces complexity and uncertainty into the environment. Given these complexities and the nuanced, evolved human sensory and cognitive capabilities, a human-in-the-loop approach becomes crucial. Here, humans use their superior sensing and decision-making abilities in complex scenarios, employing techniques such as admittance/impedance control to guide the robot in pHRC. This raises a crucial research question: How should control be distributed between humans and robots to maximise synergistic performance?

Role arbitration in pHRC, as described by [30], delves into this control distribution, seeking

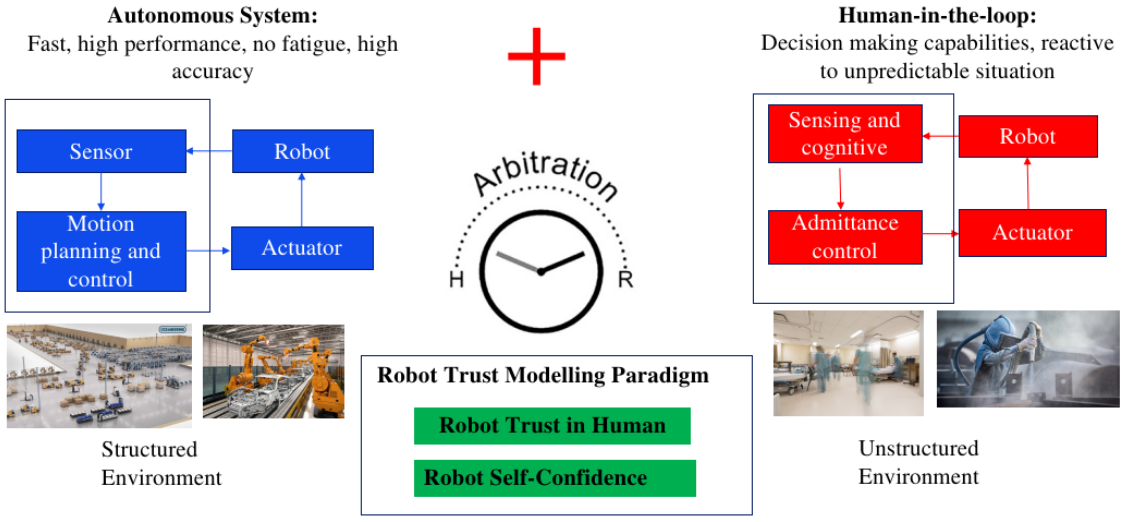


FIGURE 1.8: A role arbitration method to determine the control allocation between human and robot based on robot trust in human and robot self-confidence.

to merge the strengths of human and robot, improve performance, and minimise human physical or cognitive strain. Role arbitration refers to how the control is allocated between humans and robots for a task and allows combining the strengths of humans and robots to improve combined performance and reduce human physical or cognitive effort [30].

The role arbitration method is generally categorised based on the communication modalities. In pHRC, the use of Electroencephalograms (EEG)[31], Electromyography (EMG) [32], and Motion capture systems (MCS) [33] as external sensors brings forth notable challenges in real-world industrial settings. These challenges primarily revolve around the high cost of these sensors, making them less accessible for small enterprises and researchers. Furthermore, the need for individuals to wear these sensors or specialized suits raises issues of comfort and wearability, potentially hindering natural movement and impacting the precision of data collection. EEG and EMG sensors, in particular, are prone to low signal-to-noise ratios, diminishing measurement accuracy due to the noise generated by muscle movements during pHRC activities. While force-torque sensors and encoders are more commonly implemented in pHRC for their ability to capture physical interaction, they fall short in fully capturing the comprehensive human performance in pHRC. Therefore, a practical role arbitration method that is comprehensive enough to be applicable across various applications presents a novel research topic worthy of investigation.

The willingness of a human to rely on the robot taking control during pHRC depends on the difference between the human's trust in the robot and the human's trust in themselves to perform the task [34]. The robot can take control when human trust in the robot is higher than trusting themselves, otherwise, the human is in control. To the best of author knowledge, the role arbitration based on robot-to-human trust and robot self-confidence (TSC) is a novel research question which have not been investigated yet.

1.5 Questionaire-based Subjective Evaluation Method in HRC

Understanding the preferences of human users is critical when developing pHRC systems. Rating Scale (RS) methods, for example the NASA Task Load Index (NASA-TLX) Likert scale, is widely used to evaluate the user impressions of robotic systems [35–37]. It is simple to implement as it just requires participants to provide scores against some defined attributes or criteria on a given scale.

However, criticism of this method in relation to its limitations has been persistently suggested. Kieruj [38] has discovered that the length of response scales influences the responses. It is also addressed that the cultural factor results in differences in scale completion rates and familiarity with scales [39, 40]. Another challenge associated with Rating Scale is the risk of careless or disengaged responses. Participants who are not fond of the survey or find it tedious might rush through the questions, leading to haphazard results with diminished reliability and validity [41]. All of these limitations of the Rating Scale method lead to bias and noise in its results. The study of an alternative method instead of RS for pHRC is a novel research topic to be investigated.

1.6 Research Question

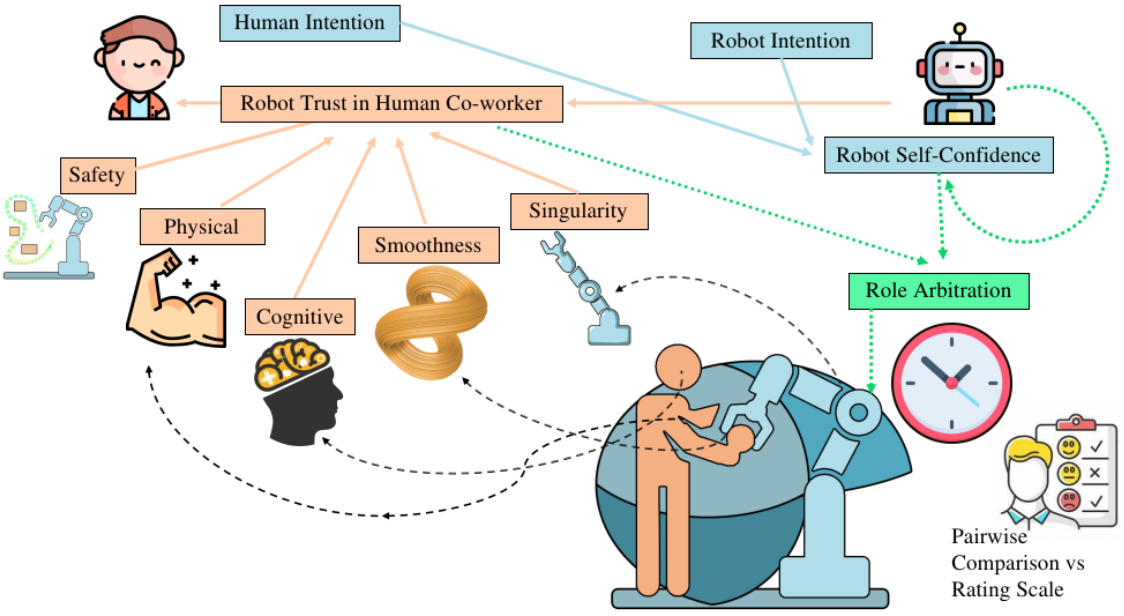


FIGURE 1.9: An abstract figure for the research. A computational model of robot trust in human co-worker which takes consider of safety, singularity, smoothness, physical and cognitive - Q1. A computational model of robot self-confidence which takes consider of human co-worker and the autonomous control's intention - Q2. And a role arbitration method which input is robot trust in human co-worker and self-confidence - Q3. And a new pairwise comparison subjective evaluation method - Q4.

On the cusp of a technological revolution, the symbiotic relationship between humans and robots is becoming more intricate than ever before. As these machines transition from mere tools to collaborative partners, it is no longer just about the tasks they perform, but also the dynamic they share with their human counterparts. A deep understanding of the trust, confidence, and arbitration mechanisms within the framework of pHRC is not just a scientific endeavour, but a social imperative. With the future poised to be shaped

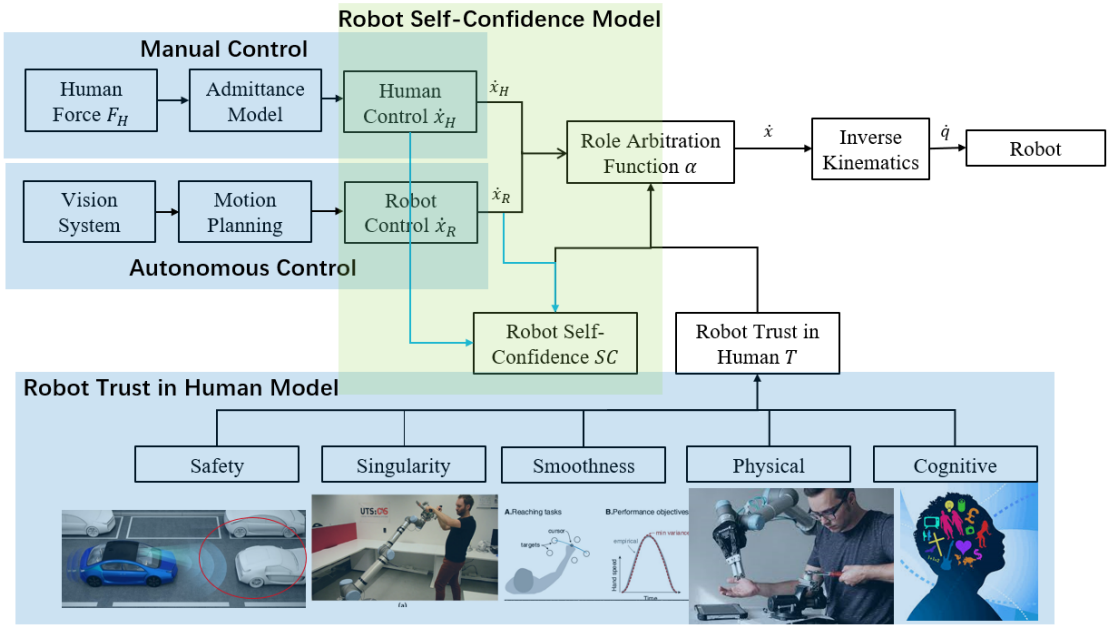


FIGURE 1.10: Control diagram of robot trust and self-confidence based role arbitration method for physical human-robot collaboration.

by this partnership, this thesis aims to address the pressing issue: How do we precisely model, measure, and enhance the intricate interplay in pHRC, specifically focussing on the nuances of robot trust in humans, robot self-awareness, and the resulting collaborative roles? Delving into this will not only optimise performance outcomes, but also pave the way for more harmonious and productive human-robot synergies in the future. Based on the motivation mentioned above, the research question is formulated as:

Q: How can trust and self-confidence be used to create effective arbitration in pHRC?

The abstract figure and control diagram are shown in Figure 1.9 and 1.10, respectively.

1.6.1 Research Sub-questions and Methodology

To elucidate the main research question, the following subquestions have been formulated:

Q1: What are the effective metrics for measuring robot trust in human collaborators within pHRC?

Q2: How can robot self-confidence be measured in pHRC?

Q3: How might we arbitrate roles based on a robot's trust in human co-workers and its self-confidence?

Q4: Which techniques are most effective and efficient to assess human subjectivity in pHRC?

1.6.1.1 Computational Model for Gauging Robot's Trust in Human Collaborators

Trust between individuals is mutual in human-human interactions. To mirror this in a human-robot pair, we introduce a computational model representing a robot's trust in its human counterpart, particularly in the realm of pHRC. This trust model is correlated with the performance metrics of human collaborators. Factors that encapsulate this performance in pHRC include safety considerations, robot singularity, task smoothness, physical ability, and cognitive state. Validation of this trust model is achieved through three experiments employing a collaborative robot. A pivotal contribution of this study is crafting an intricate, objective, and holistic computational model to gauge a robot's trust in human collaborators within the sphere of pHRC.

1.6.1.2 Computational Model of Robot's Self-Confidence

A robot's sense of self-confidence is sculpted by human acknowledgement and endorsement of its control decisions. When human partners concur with robot control, the robot's self-confidence surges, and the opposite holds true. This framework is adept at navigating challenges in unstructured environments or tasks, be it alterations in a task by a human collaborator, interference from sensor noise, or unforeseen hindrances, leading the robot to either falter in completing a predefined task or deviate from anticipated behaviour. The robot's self-confidence tends to wane in unpredictable settings, especially when there is a chasm between human intent and the robot's misguided directives.

1.6.1.3 Role Arbitration Informed by Robot Trust and Self-Confidence

The proclivity of human collaborators to lean on robot autonomy depends on juxtaposing their trust in the robot against their own self-confidence. Should a human collaborator have trust in a robot's competence to autonomously complete a task, they will be more inclined to delegate the task to the robot. In contrast, a stronger self-belief makes humans lean more on their own capabilities, overriding robot autonomy. Thus, the extent of human reliance on robot autonomy is tied to the balance of trust they place in the robot and their self-assuredness. Within the human-robot role arbitration paradigm, humans and autonomous systems collaboratively steer robot actions during tasks, amalgamating the strengths of both entities to bolster combined efficiency and alleviate human strain, be it

physical or cognitive. This arbitration in human-robot collaboration is dynamic, influenced by numerous factors, including physical exertion, environmental changes, and trust levels. To address this fluidity, this research takes a trust-centric arbitration approach. This thesis proposes a computational model that encapsulates both trust and self-confidence, termed the TSC, within pHRC. A consequent arbitration method grounded in the TSC model is introduced. Empirical "human-in-the-loop" experiments with a collaborative robot validate the efficacy of this TSC-informed arbitration approach. Preliminary findings illuminate that our method not only accentuates synergistic human-robot performance, but also curtails the workload on human collaborators, whilst enhancing their subjective experience.

1.6.1.4 Effective Techniques for Gauging Human Subjectivity in pHRC

In pHRC research, qualitative tools, such as questionnaires, are widely used to capture subjective human evaluations of robotic counterparts. The likert scale assessment has garnered favour not only in pHRC but also in a wide range of HRI domains. However, singular ratings assigned to experimental conditions can inadvertently induce biases and noise. When pairs of different experimental conditions are compared rather than individually, it is possible to mitigate these negative effects, thus increasing the reliability and accuracy of the results. This has led us to postulate that PC might be a more streamlined method for pHRC studies over conventional lkert scale assessment. The PC technique ranks multiple alternatives through a succession of binary comparisons. In this discourse, we juxtapose the results from the widely recognised NASA TLX against the PC approach in two distinct pHRC trials with identical experimental conditions and a collaborative robot. One distinguishing factor is the comprehensive versus partial conditions presented to the participants. Drawing comparisons using the ANOVA, the PC technique emerges as both resilient and precise, registering a reduction 44% in experimental timelines.

1.7 Thesis Outline

This thesis undertakes a comprehensive examination of human-robot collaboration, with a specific emphasis on trust, self-confidence, and the role arbitration method. To ensure clarity and methodical progression, the study is organised into distinct chapters, each dedicated to a particular aspect of the central research theme. Subsequently, the structure of the thesis is delineated.

1.7.1 Chapter 2

This chapter is a review of related work and foundational knowledge, introducing the reader to the context and significance of pHRC. Here, this thesis explores the relevant literature on trust in HRI, delves into role arbitration methodologies, and examines the varied techniques employed to gauge subjectivity in pHRC.

1.7.2 Chapter 3

In Chapter 3, the focus shifts to the process of creating a computational model to capture a robot's trust in its human collaborator within the realm of pHRC. The development of this model is evaluated using empirical data from three experiments involving a collaborative robot to validate the proposed trust model.

1.7.3 Chapter 4

The fourth chapter expands on the trust model of Chapter 3 by introducing the TSC. This comprehensive model integrates the concepts from Chapter 3 with a new layer: the robot's self-confidence. This is incorporated into a role arbitration method grounded in the TSC framework. The validity of this method is examined through human-centric experiments with a collaborative robot.

1.7.4 Chapter 5

Chapter 5 pioneers a novel methodological approach to subjective measurement. Here, this chapter evaluates the use of pairwise comparison for pHRC experiments, juxtaposing it with traditional rating scales, exemplified by tools like NASA-TLX.

1.7.5 Chapter 6

This chapter presents an extensive experimental evaluation of the role arbitration method that is based on robot trust and self-confidence. A rigorous statistical analysis is used to draw on objective and subjective evaluation criteria. Additionally, the physical and cognitive performance models, initially presented in Chapter 3, are revisited and validated within real-world pHRC scenarios.

1.7.6 Chapter 7

This concluding chapter consolidates the findings and insights derived from the study. The implications of the research are examined, leading to definitive conclusions. Looking ahead, potential directions for future research in this domain are suggested.

1.8 Publication

The work presented in this thesis has been presented in the publications listed below. Qiao Wang is the primary contributor and the first author in these publications.

1. Qiao Wang, Dikai Liu, Marc G. Carmichael, Stefano Aldini, Chin-Teng Lin. Computational Model of Robot Trust in Human Co-Worker for Physical Human-Robot Collaboration. *IEEE Robotics and Automation Letters (RAL)*, pp. 3146–3153, 2022

2. Qiao Wang, Dikai Liu, Marc Carmichael, Chin-Teng Lin Lin. Robot Trust and Self-Confidence Based Role Arbitration Method for Physical Human-Robot Collaboration. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023

3. Qiao Wang, Ziqi Wang, Marc Carmichael, Dikai Liu, Chin-Teng Lin. Comparison of Rating Scale and Pairwise Comparison Methods for Measuring Human Co-worker Subjective Impression of Robot during Physical Human-Robot Collaboration. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024

4. Qiao Wang, Ziqi Wang, Marc Carmichael, Dikai Liu, Chin-Teng Lin. Experimental Evaluation of Robot Trust and self-confidence Based Role Arbitration Method for Physical Human-Robot Collaboration. *2024 IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2024 (Manuscript to be submitted)

Chapter 2

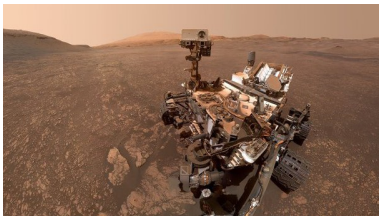
Review of Related Work

HRI is related to the multidisciplinary study of the dynamics, design, and evaluation of robot systems intended for human use or collaboration [42]. Central to HRI is the notion of interaction, which inherently demands a communicative nexus between humans and robots. The nature of this communication is significantly dependent on the spatial proximity between the interacting parties. Consequently, the interactions can be broadly classified into:

1. **Remote interaction:** The human and the robot are not in the same location and can be distant in terms of space or time [42].
2. **Proximate interaction:** The human and the robot are in the same location in terms of space or time [42].

In the realm of remote interactions, several applications can be delineated. Telemanipulation is exemplified by systems like the Da Vinci robot (Figure 2.1b), while supervisory control (Figure 2.1a) can be seen in scenarios such as the remote operation of the Mars Rover, which involves spatial and temporal separations. For proximate interactions, the spectrum encompasses physical human-robot collaboration, as previously elaborated, and extends to social interactions. In these instances, robots and humans engage on a peer-like level, epitomised by platforms such as Nao, which fosters emotional and social exchanges with humans (Figure 2.1c).

This chapter reviews previous work related to the contribution of this thesis. Section 2.1 provides a general overview of the state-of-the-art in pHRC and collaborative robotics. Section 2.2 provides a general overview of trust in HRI. Section 2.3 provides a general overview of the role arbitration method in the HRI. Section 2.4 provides a general overview of the subjective evaluation method in the HRI.



(a) Example of remote interaction: supervisory control (Mars Rover)

(b) Example of remote interaction: telemanipulation (Da vinci surgical robot)

(c) Example of proximate interaction: social interaction (Nao robot)

FIGURE 2.1: Applications and examples of remote and proximate interactions: (a) Supervisory control (Mars Rover) (b) Telemanipulation (Da vinci surgical robot) (c) Social interaction (Nao robot)

2.1 Physical Human-Robot Collaboration

2.1.1 Physical Human-Robot Interaction

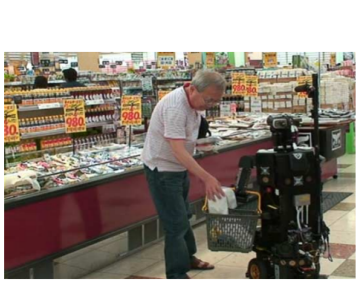
pHRI refers to scenarios in which humans and robots are in direct contact or share a workspace, requiring physical interactions [43]. It emphasises the importance of safety, intuitiveness, and efficiency when humans and robots collaborate in a physical manner [44]. pHRI encompasses the development of robotic systems that are sensitive to human presence, can adjust their actions in real-time based on human movement or input, and are designed to operate in shared environments without causing harm or discomfort to humans [45]. This field addresses both the technical challenges of robot design, sensing, and control, as well as aspects of human factors, such as user trust, understanding, and acceptance of robotic systems [46]. In this section, the classification of pHRI will be introduced. Classification depends mainly on the frequency of physical interaction and the level of autonomy of the robot, and can be classified into three types: supportive, cooperative and collaborative [47].

1. **Supportive Interactions:** In this category, the robot acts more as an assistant, enhancing the human's performance rather than being essential to the task itself. Such interactions usually involve the robot offering tools, information, or materials to humans. Examples include shopping assistants [48] for the elderly, where the robot carries a basket for the elderly as shown in Figure 2.2a. Safety is paramount, with an eye on avoiding unwanted contacts or collisions. Physical interactions are rare and often brief. Effective human-robot communication, especially bilateral gesture cues,

is critical for tasks like turn-taking and close-proximity activities instead of frequent physical contact.

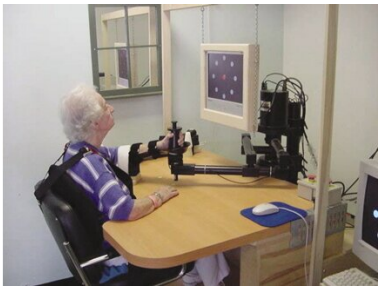
2. **Cooperative Interactions:** Here, both humans and robots actively participate in the task, dividing labour based on their respective strengths such as handover of a bottle by human and robot [49] as shown in Figure 2.2b [49]. They interact more often, either by passing parts/tools or through haptic cues that switch the robot's behaviour. Humans handle tasks that require dexterity or decision-making, while robots handle repetitive tasks or those that require precision. Although physical space is shared more than supportive interactions, physical exchanges are still mostly transactional.

3. **Collaborative Interactions:** This level of interaction involves more direct and continuous physical collaboration between the human and the robot. In the two preceding interaction types, the robot assumes a more passive role. However, in collaborative mode, the robot adopts a more active and independent stance, equipped with self-intelligence. This shift underscores the robot's capability to engage more dynamically and autonomously in the collaborative process. In this level of interaction, the robot can work in direct or indirect contact with humans, either through physical touch or through a shared object. These interactions cover tasks such as cooperative lifting, kinaesthetic teaching, handling flexible materials, and rehabilitation therapy [3] (Figure 2.2c).



(a) Example of Supportive Interaction: shopping assistant robots for aiding seniors [48]

(b) Example of Cooperative Interaction: human and robot hand-over a bottle. [49]



(c) Example of Collaborative Interaction: rehabilitation robot for stroke patient [3].

FIGURE 2.2: Examples of classification of physical human-robot interaction (a) Supportive Interaction (b) Cooperative Interaction (c) Collaborative Interaction.

In essence, the categories range from robots providing supportive roles to more direct, shared-task engagements, each with increasing degrees of physical interaction and collaboration. However, in current research, the supportive and cooperative types of interactions

are well-researched. With advances in autonomous technology, robots can now function as independent entities. This has paved the way for a collaborative paradigm between humans and robots. Given this promising trajectory, the primary focus of this thesis is on physical human-robot collaboration.

2.1.2 Collaborative Robot

Cobots are the technological embodiment designed for pHRC scenarios. Although pHRC describes the nature and dynamics of the interaction, cobots are the tools that facilitate this interaction. In the context of pHRC, cobots are developed to be inherently safe, making it possible for humans and robots to collaborate and share tasks seamlessly in a physical workspace. In summary, cobots serve as the primary enablers of effective and safe pHRC in various applications, from manufacturing to healthcare.

Traditional industrial robots are generally operated behind fences and away from humans due to the potential concern for safety and perform repetitive and labour intensive tasks [50], as shown in Figure 2.3a. With the advancement of robotics research (such as hardware and software), especially collaborative robots [9], cobots can now actively and safely interact with human co-workers in the same workspace. Examples of general purpose cobots, shown in Figures 2.3(b-d) include; KUKA Robots (KUKA LBR iiwa) [51], Universal Robots (UR 10) [52], and Rethink Robotics (Sawyer) [53]. Examples of cobots for specific applications include; the ANBOT for industrial applications (Figure 1.4), smart hoist robot (Figure 1.5) and rehabilitation robots (Figure 1.2.4) in medicare applications.

(a) Traditional Industrial Robots (b) KUKA LBR iiwa (c) Universal Robot 10 (d) Rethink Robotics Sawyer

FIGURE 2.3: Examples of traditional industrial robot (a) and collaborative robots (b)(c)(d)

There are many considerations when it comes to designing cobots that coexist with human co-workers safely [54], which need to satisfy the ISO/TS 15066 standards for cobots [55]:

1. Compliant actuators:

(a) Series Elastic Actuators [56]: In the realm of advanced actuation systems, Series Elastic Actuators (SEAs) have emerged as a notable innovation. Characterised by the integration of an elastic component, typically a spring, in series between the actuator's output and the load, SEAs offer multifaceted benefits. Not only do they enable precise force measurements through the deflection of the elastic element, but they also enhance shock tolerance and safety, particularly in human-robot interactions. The inherent compliance of SEAs protects against abrupt impacts, while simultaneously expanding control possibilities. Their introduction represents a significant step towards creating more adaptable and safe robotic systems, especially in dynamic environments.

(b) Series Damper Actuators (SDA) [57]: A standard SDA setup encompasses a control module paired with two core hardware modules: a motor, which may or may not incorporate a gear transmission, and a strategically integrated viscous damper. This configuration is meticulously crafted to exercise precise control over the relative velocity within the damper, subsequently achieving the intended force based on the designated damping coefficient.

When contrasted against the Series Elastic Actuator (SEA) system, which relies primarily on an elastic element, the SDA distinguishes itself by employing a series damping component. This key alteration paves the way for a system order reduction of one. The practical implications of this are quite profound: the SDA might possess a wider operational bandwidth compared to its SEA counterpart. A notable merit of the SDA framework lies in its inherent flexibility. By leveraging a suitable damper design, the damping coefficient can be dynamically modulated in alignment with changing environmental conditions. Such adaptability finds practical applications in scenarios necessitating variable forces. For example, the system can seamlessly adjust the damping coefficient in response to varying force magnitudes, thus optimising the relative velocities within the damper. This feature ensures that the SDA delivers unparalleled force fidelity across a spectrum of force ranges.

Another commendable characteristic of the SDA is its intrinsic ability to absorb impacts. Due to the energy dissipation properties of the series damper, the system is well-equipped to protect walking robots, haptic interfaces, and robotic manipulators from potential damage upon external impact exposure.

Yet, no system is devoid of challenges. The very attribute that equips the SDA with impact absorption, its energy dissipation property, simultaneously curtails its operational efficiency. In a comparative context, the SEA, with its energy-conserving elastic element, has the potential to offer superior efficiency.

However, questions persist about the real-world efficacy of this energy conservation, especially when SEAs are integrated into complex robotic systems like legged robots.

(c) Variable Impedance Actuators [58]: Variable Impedance Actuators (VIA) have gained the attention of researchers due to their suitability for dynamic interactions with unpredictable environments, including human involvement, where traditional stiff actuators fall short. Variable impedance actuators (VIAs) are a type of actuator system designed to adjust their mechanical impedance, which encompasses both stiffness and damping properties. Unlike traditional actuators, which typically have fixed or "stiff" impedance characteristics, VIAs can modulate their impedance based on specific requirements or environmental conditions.

2. Speed and Separation Monitoring: The Speed and Separation Monitoring (SSM) technique is used to prevent collisions by constantly monitoring human movement within the robot workspace [59]. This approach calculates the space needed to bring a robot to a safe state, such as stopping, ensuring a minimal protective distance that aligns with its deceleration or slowing pace.

3. Power and Force Limitation (PFL) [60]: The "Power and Force Limitation (PFL)" method is designed to minimise the adverse effects of unintended impacts, ensuring they don't cause harm. Although the foundational principle of safe human-robot interaction posits that a user should never sustain an injury, sporadic accidental contacts might result in a sensation akin to minor pain experienced in daily activities. Regulatory guidelines for the PFL technique establish criteria to ensure that all contacts remain within such boundaries, based on research on the impacts on the human body. Importantly, these studies never induce serious injury. In collaborative robotics, the goal is to assess levels of physical interaction far from causing injury, in line with the objective of increasing human-robot interactions in shared spaces. A vital metric for PFL, although challenging to measure and validate, is the pressure exerted on the human body. It is essential to assess the force's magnitude and duration from a robot system on the body's impacted surface and consider strategies to reduce it or control actions when there is potential significant risk.

2.2 Trust in Human-Robot Interaction

Social Human-Robot Interaction (sHRI) refers to the study and practice of designing, implementing, and evaluating robots that are capable of engaging with humans in socially meaningful ways. This field encompasses the development of robotic systems that can

understand and adhere to social norms and cues, communicate effectively through both verbal and non-verbal means, and adapt to the complex dynamics of human social environments. Social HRI aims to create robots that can participate in a variety of roles, from companions and assistants to collaborators and autonomous agents, in settings ranging from homes and workplaces to public spaces and educational environments. The goal is to enhance human experiences, improve task efficiency, and foster positive relationships between humans and robots by ensuring that interactions are intuitive, natural, and satisfying for the human participants [26].

In sHRI, numerous variables influence the dynamics and outcomes of interactions, such as Robot Design and Appearance, Behavior and Autonomy, Reliability and Trust [26].

Robots and advanced autonomous systems promise significant advantages by helping humans perform various tasks. However, the full potential of these advantages is occasionally not realised because of inappropriate human interactions. In interactions with automation, humans sometimes neglect beneficial system features. This phenomenon, termed "disuse" or "under-reliance" [61], indicates a reluctance to employ automation even when it is beneficial. Furthermore, there are instances in which individuals improperly oversee automated systems, such as removing alarms or blindly following their directives even if inappropriate [62]. A related concern is "automation bias" [63], where users allocate undue credibility to automated recommendations over other inputs, including human judgment. When these automated suggestions fail, the results can be disastrous [64], leading to neglect of crucial situations or misguided actions.

Both inexperienced and experienced users exhibit these inclinations. For example, in [65], even experts in the domain displayed excessive trust in diagnostic systems [65]. The Aviation Safety Reporting System has numerous accounts from pilots attributed to oversight errors to overreliance on automated systems such as autopilots[66]. On the contrary, when organisational rules enforce distrusted automation, operators might circumvent it, essentially leading to its disuse [66].

Research [67] highlights the impact of trust on automation reliance; people are inclined to depend on automation they trust and abandon what they distrust. Trust plays a pivotal role in decisions about automation supervision and utilisation [68]. A recurring theme in the automation trust literature associates complacency with over-reliance, insufficient monitoring, and reduced alertness to automation [68]. Optimal performance in human-automation collaboration requires calibrated trust. Both under-utilisation and misappropriation of automation, resulting from miscalibrated trust, have been linked to mishaps [69].

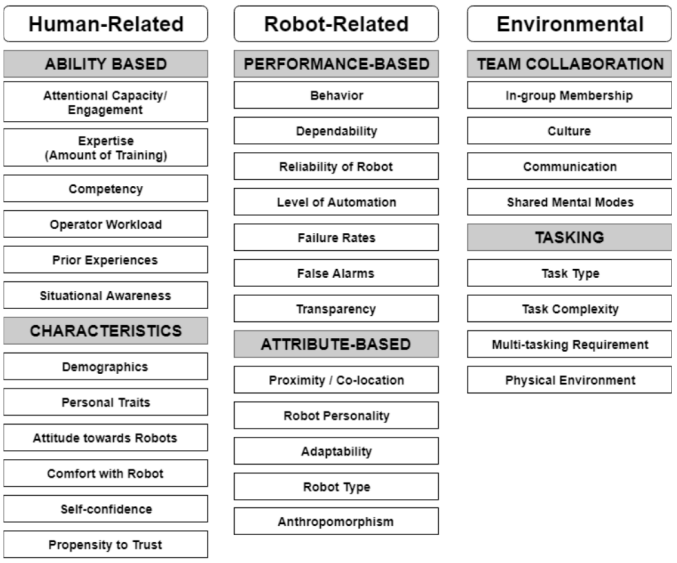
Trust is a multifaceted concept that has attracted attention from researchers across numerous disciplines, ranging from social psychology and human factors to industrial organisation. The breadth of this attention has been both a boon and a challenge; while it has enriched our understanding of trust from various perspectives, it has also resulted in many definitions, theories, and applications [70].

Given the myriad contexts in which trust is examined, its definition has been adapted to fit diverse situations. In some studies, trust is delineated as an attitude, while others perceive it as an intention or a behaviour. This variety in conceptualisation has inhibited the establishment of a universally accepted definition, even within specific domains such as interpersonal relations and human-automation interactions [71]. However, a consensus has emerged that perceives trust as a multidimensional psychological position, which encompasses beliefs and anticipations about the trustworthiness of a trustee. This position is forged from experiences and engagements with the trustee, particularly in scenarios typified by uncertainty and potential risk [72].

Deeper analysis of trust reveals that it has both cognitive and affective elements. Within interpersonal dynamics, the affective dimension of trust is evident, as it requires perceiving others as intrinsically driven by care and consideration to protect the interests of the person who places trust [73]. When it comes to automation, the cognitive dimension takes precedence. Here, trustworthiness revolves primarily around the anticipation that automation will perform its designed function [74].

2.2.1 Trust Modelling in Human-Robot Interaction

Trust is key to achieving a successful HRI [27]. From [27], trust can be defined as "attitude that an agent (automation or another person) will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability". The human may disuse (under-reliance), or misuse (over-reliance) the robot without an appropriate level of trust [28]. Figure 2.4a shows the factors that affect trust in HRI, including human-related, robot-related, and environmental-related [75]. Since pHRC falls within the broader scope of HRI, the factors that influence trust in HRI similarly impact trust in pHRC. which is shown in Figure 2.4b. The highest correlation has been shown between robot performance (such as dependability, reliability, and failure rates) and the evolution of trust [75], [76]. Therefore, computational trust models based on robot performance have been well-researched by several research groups, which will be introduced in detail in the following section.



(a) Meta analysis of the factors Affecting Trust in Human-Robot Interaction [75]

Element	Trust-related themes	Lower level theme	Frequency
Robot	Performance	Motion of the robot	21
		Robot and gripping mechanism reliability	20
	Physical attributes	Robot's size	18
		Robot's appearance	15
Human	Safety	Personal safety	17
		Robot's safety features	11
		Safe programming of the robot	6
	Experience	Priorexperiences with robots	14
		Mental models of robots	9
External	Task	Task complexity	15

(b) Meta analysis of the factors Affecting Trust in physical Human-Robot collaboration [76].

FIGURE 2.4: Meta analysis of the factors Affecting Trust in (a) Human-Robot Interaction (b) Physical Human-Robot Collaboration.

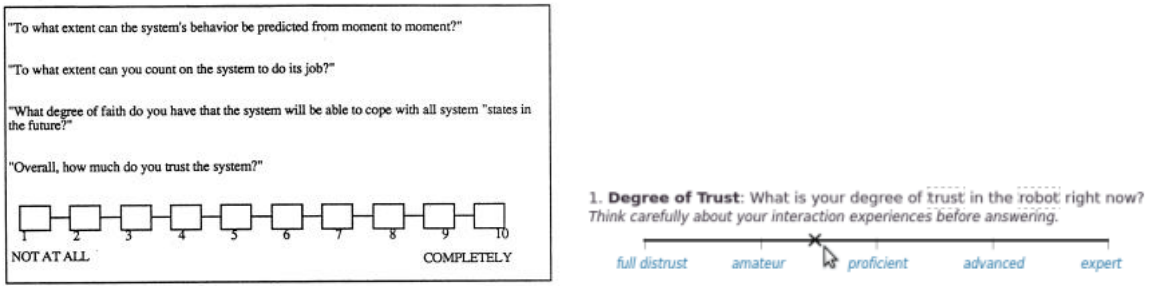
Trust in HRI has been extensively researched. A common method for measuring human-to-robot trust is through subjective questionnaires, which researchers provide during the experiment [29]. However, researchers can only acquire the questionnaire results after the experiment is completed, or interrupt the experiment to obtain the results while the experiment is being performed. The results can therefore not be directly used for real-time computational modelling and controlling the robot. Therefore, real-time computational trust models must be developed so that robots could adapt their behaviour in real-time based on trust to improve human-robot joint performance and reduce human co-worker physical and cognitive efforts.

The first research on the computational trust model of human trust in robots comes from Lee and Moray [5]. This research found that the occurrence of faults and performance are the two most significant factors affecting the variation of trust. After each trial, the Likert questions with a maximum possible score of 10, meaning complete trust in the robot, used to evaluate the human co-worker trust in the robot is shown as: "To what extent can the system's behaviour be predicted from moment to moment?", "To what extent can you count on the system to do its job?", "What degree of faith do you have that the system will be able to cope with all system states in the future?" and "Overall, how much do you trust the system?" as shown in Figure 2.5a. The researchers developed a dynamical trust model which employs an autoregressive moving average model based

on operator's subjective rating scales of trust mentioned above, the current and previous automatic controller's performances, and faults.

Xu [6] proposed an online probabilistic trust inference model which employs a dynamic Bayesian network to estimate trust based on the history of interactive experience. This trust model could maintain belief distributions for these performance centric trust measures, based on various factors of the interaction experience. This probabilistic representation is useful for inferring the human expected trust state at a given time, as well as the amount of uncertainty of each such estimate. The subjective trust questionaire is obtained to train the trust model and evaluate the prediction accuracy. The post session trust feedback questionaire is shown as Degree of Trust: "What is your degree of trust in the robot right now" as shown in Figure 2.5b, the scale is listed as a modified Visual Analog Scale (VAS) full distrust, amateur, proficient, advanced and expert.

For existing human-to-robot trust models [5, 6], subjective questionnaires about trust are used to determine the parameters used in the model. However, this approach cannot be performed for the development of robot-to-human trust models, as robots do not have an intrinsic model of trust to be queried. Instead, a model of trust needs to be developed based on other measures.



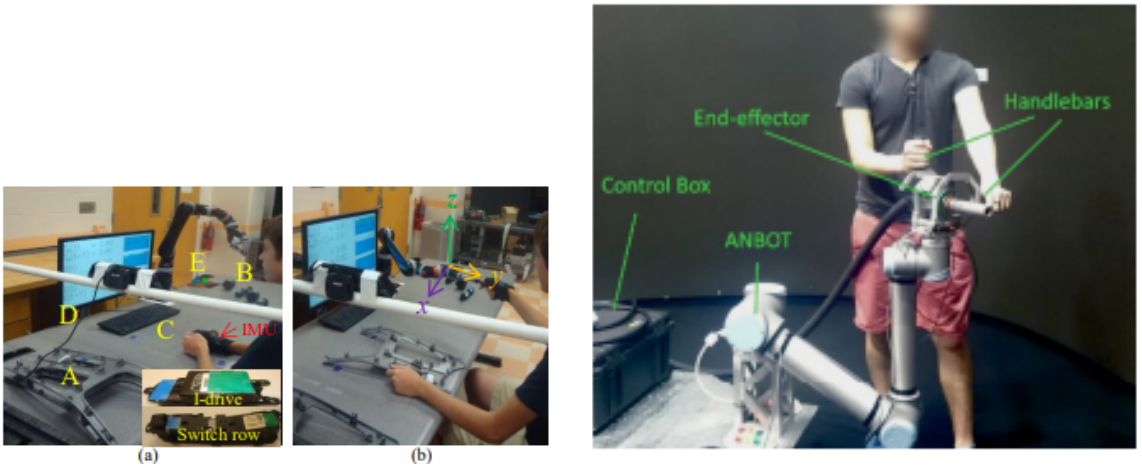
(a) Post-session trust questionnaire [5] (b) Post-session trust questionnaire [6].

FIGURE 2.5: Post-session trust questionnaire (a) autoregressive moving average model [5] (b) online probabilistic trust inference model. [6]

Trust is inherently bidirectional between the human co-worker and the robot co-worker during the interaction. In addition to models of human trust in robots, models of robot trust in humans were also studied by several researchers. Rahman [7] proposed a robot-to-human trust model based on for handover motion planning tasks. The robot adapts its handover configuration and motion by exploiting kinematic redundancy, depending on robot trust in humans. The real time trust measurement is modelled as a function of human performance. The human performance modelling is based on normalized values for human hand speed for part manipulation, and part gripping and releasing for manual manipulation and part attachment during assembly. The speed is measured through IMU.

Also, the human fault modelling is based on the measurement of orientation of assembled parts. A Kinect camera is used to identify the orientation of parts through the OpenCV library.

Tran [8] proposed a robot-to-human confidence model based on the fluid-stochastic Petri net model for grit-blasting tasks in pHRC. The confidence model is a function of human performance, which is a function of application-specific measures including blasting angle, variation in blasting path and blasting angle.



(a) Robot to human trust model for assembly task [7]

(b) Robot to human trust model for grit-blasting task [8].

FIGURE 2.6: (a) Collaborative assembly in hybrid cell and handover configuration and the axes of end-effector [7] (b) ANBOT equipped with nozzle and hose mounted for cooperative grit-blasting [8]

It is notable that existing robot-to-human trust models [7, 8, 77] have utilised performance measurements for their development. However, these have been restricted to measurements specific to the particular task of interest.

The proposed approach in this thesis directly aims to address the aforementioned limitation by using performance measures that are less specific to the task. The robot-to-human trust model takes into account many pHRC factors, including safety, robot singularity, smoothness, physical performance, and cognitive performance, which evaluates human co-worker performance more comprehensively.

2.3 Role Arbitration Method in physical Human-Robot Collaboration

Historically, robots were designed to passively adhere to human directives, using the cognitive supremacy of humans. In the realm of pHRC, robots are designed to respond passively and compliantly to human-induced interaction forces, often through mechanisms such as admittance or impedance control [78]. As advances in robotics burgeoned, particularly with the incorporation of sophisticated sensors and intricate control systems, robots began to manifest increased intelligence. Consequently, this bolstered the robot's level of autonomy, enabling it to assume more proactive roles in tasks and potentially alleviate the cognitive and physical strains on the human collaborator. Such enhancements not only empower robots to spearhead certain tasks but also enable them to negotiate intentions or instigate role alterations with their human counterparts.

Yet, the extent of robot 'intelligence' remains constrained, often faltering in the face of limited data about the task or environment or other unforeseen scenarios. Such instances necessitate human intervention, especially when confronting unstructured challenges that remain outside the robot's pre-programmed purview. As a result, discerning the appropriate moments to transition between *leadership* and *followership* roles for humans and robots becomes a critical research question.

In the context of human-robot role arbitration, control over a robotic task is bifurcated between a human collaborator and an autonomous controller. The dichotomy of the leader and follower roles emerges as a prevalent paradigm in this scenario [30]. The underlying rationale for human-robot role arbitration hinges on amalgamating the unique capabilities of both entities: the strategic planning and cognitive prowess of the human collaborator juxtaposed against the robot's endurance and strength capabilities. This synergy aims to optimise the joint performance of human-robot while simultaneously attenuating the cognitive and physical demands on the human participant [79]. Such arbitration finds relevance in diverse domains, from robotic rehabilitation [80] to teleoperation [81].

Figure 2.7 shows a conceptual depiction that elucidates the dynamics of role arbitration in the context of pHRC, as introduced by Losey et al. [30]. Central to this framework is the principle of arbitration, which essentially determines the proportion of control apportioned to either the human collaborator or the robot, contingent upon the prevailing intents emanating from either entity. As shown in Figure 2.7, as the transition from left to right, the leftmost segment illustrates human intent, denoted as u_h. This is typically interpreted as either a velocity or force control command derived from the principles of admittance or impedance control. Centrally, we have the robot's intent, represented as u_r, which is translated as the velocity and position control commands directed towards the robotic

manipulator. These commands are governed by the autonomous system's motion planning and control mechanisms. The rightmost section illustrates the resultant control command, u, which emerges as a synthesis of human and autonomous inputs, achieved through the role arbitration methodology.

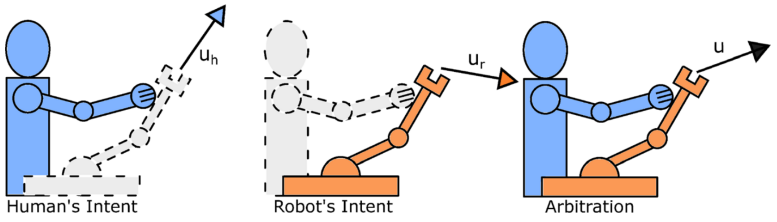


FIGURE 2.7: Conceptualised framework of the role arbitration method in physical human-robot collaboration [30].

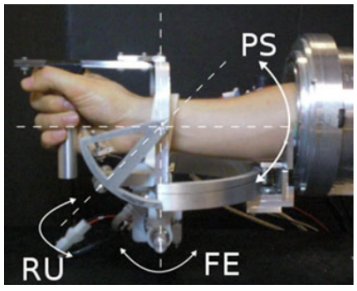
The role arbitration method is generally categorized based on the communication modalities. Those are the common sensors and information used in decision-making for the dynamic of role arbitration:

1. **Force-Torque Sensor:** Due to the constant physical couplings between humans and robots during pHRC, the force-torque sensor is a critical sensor used in the applications of pHRC. The time-varying force or torque information obtained through a force-torque sensor can be interpreted as the intention and capability of the human being. Pehlivan [82] identifies the ability of the stroke partient in real-time through force-torque information and provides assistance as needed through an upper limb rehabilitation robot, as shown in Figure 2.8a.

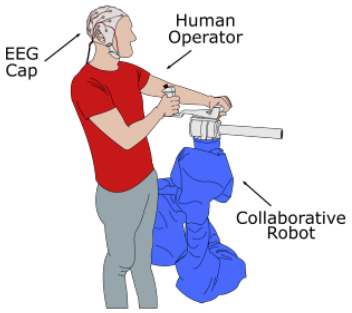
2. **Electroencephalogram (EEG):** EEG could help identify the desired movement through motor imagery or cognitive information, such as cognitive conflict or cognitive workload [83, 85]. The method has been implemented into the singularity avoidance method to optimise the predictability of the human co-worker during pHRC as shown in Figure 2.8b [31].

3. **Electromyography (EMG):** Due to constant physical engagement during pHRC, muscle fatigue is an important factor to consider. Therefore, EMG is a popular choice for determining fatigue level and deciding the arbitration role. Peternel [32] applies EMG to measure fatigue level during the co-manipulation of the sawing task. When the fatigue level drops to a certain threshold, the robot takes control as shown in Figure 2.8c.

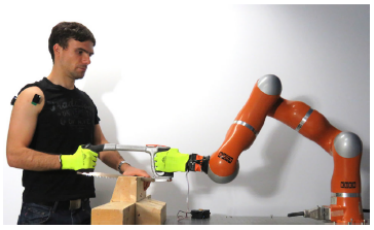
4. **Kinematic Information of Human Body (Motion Capture System):** Carmichael [33] proposed a method to estimate the physical assistance needed by using a musculoskeletal model using the motion capture system (MCS). When the task demanded



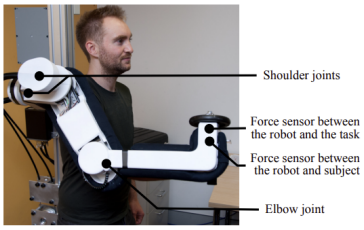
(a) Force-torque sensor [82].



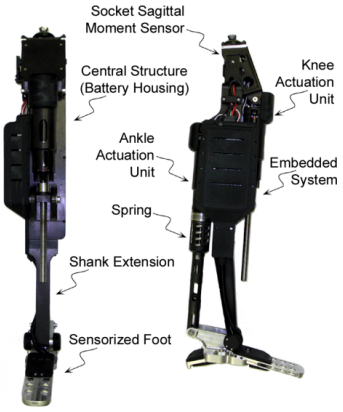
(b) Electroencephalogram (EEG) [83].



(c) Electromyography (EMG) [32].



(d) Kinematic Information of Human Body (Motion Capture System) [84].



(e) Kinematic Information of Robot (Encoder)

FIGURE 2.8: The role arbitration method is generally categorised based on the communication modalities. (a) Force torque sensor (b) Electroencephalogram (EEG). (c)Electromyography (EMG). (d) Kinematic Information of the Human Body (Motion Capture System). (e) Kinematic information of the robot (encoder).

force is lower than the available physical strength, assistance as needed will be provided to the human co-worker in the rehabilitation robot application. The proposed model has been validated through admission control [84], as shown in Figure 2.8d.

5. **Kinematic Information of Robot (Encoder):** Kinematic information from the robotic system, such as end-effector position, velocity, acceleration, and joint position, velocity, acceleration, can be important information for inferring human state or intentions due to coordinated motor movement between human and robot. In the study in [86], a control structure is presented paired with a method to recognise user intent, to facilitate real-time supervision control of a powered lower limb prosthesis. The methodology discerns user actions such as standing, sitting, or walking by analysing patterns discerned from sensor data from the prosthesis, notably without

requiring instrumentation of the unaffected leg. The intention detection mechanism employs temporal features derived from prosthesis signals, which are then dimensionally reduced to improve computational efficiency. Subsequent to this reduction, these signals inform the training of models that classify identified patterns into distinct actions such as standing, sitting, or walking as shown in Figure 2.8e.

EEG, EMG, and MCS, as external sensors employed during physical human-robot collaboration, present significant challenges in real industrial applications. Drawbacks of these sensors include their cost, which can be prohibitive for small businesses and researchers, and issues of wearability and comfort. Systems that necessitate wearing sensors or suits can be intrusive and uncomfortable over extended periods, potentially affecting the wearer's natural movement and, consequently, the accuracy of the data collected. Additionally, EEG and EMG sensors suffer from low sensor-noise levels, leading to poor measurement accuracy, especially during physical human-robot collaboration where muscle movement constantly introduces noise into the signals. While force-torque sensors and encoders are commonly used in physical human-robot collaboration, their limited measurement dimensions fail to comprehensively represent human performance modeling. The robot-to-human trust model considers various pHRC factors, including safety, robot singularity, smoothness, physical performance, and cognitive performance, offering a more comprehensive evaluation of human coworker performance.

2.4 Method in measuring subjective impression during Physical Human-Robot Collaboration

In HRI, subjective evaluation is a critical factor to consider. Because the human will degrade the performance and is likely not to use the robot when the subjective impression is too low [27] when using the robot. In terms of human-centred robotic design, the questionnaire is the most common and recognisable method for evaluating subjective impression during HRI. The questionnaires consist of a set of questions given to participants to assess the participant's subjective outcomes of an HRI experiment.

2.4.1 Rating Scale Method

Rating scale assessment is widely used by robotic researchers to explore the views of participants regarding robots' appearance, interaction experience, and overall satisfaction[87]. Commonly used rating scale questionnaires include NASA Task Load Index (TLX) [88], Negative Attitude toward Robot Scale (NARS) [89], and Godspeed Questionnaire [90]. All

of these questionnaires are based on a rating scale that requires participants to provide ratings against the defined criteria.

2.4.2 Types of Questionaires

2.4.2.1 Negative Attitude towards Robot Scale (NARS)

NARS is a method that reveals any pre-existing negative feelings toward robots [91] by asking participants to provide a grade between 1 and 5 for 18 items. A typical *Likert* scale was used and has the meanings of; 1 - I strongly disagree, 2 - I disagree, 3 - It is not decidable, 4 - I agree, 5 - I strongly agree. The total score reflects how much anxiety or negative feelings a human user has against robots [89], with a higher score reflecting negative feedback.

2.4.2.2 Godspeed Questionnaire

The Godspeed Questionnaire is a similar tool to NARS that assesses the participant's pre-existing attitude towards a robot. Participant's perceived anthropomorphism, animation, likeability, intelligence, and safety of a robot are assessed on a 5 point scale where 1 means Strong and 5 means Weak [90].

2.4.2.3 NASA-TLX

In this thesis, NASA-TLX is the questionnaire used to evaluate the subjective impression of the participant when operating the robot. NASA-TLX evaluates subjective workload from six dimensions: *Mental Demand, Physical Demand, Temporal Demand, Performance, Effort* and *Frustration*. A 21-point scale is used, where 0 means Very Low (workload) and 21 means Very High (workload). By calculating the weighted average of the ratings on these six subscales, the overall workload of the task can be derived [88]. NASA TLX was applied in [36] and [92] where in the former the cognitive workload during a heavy object manipulation task and the user's satisfaction with the proposed control schemes was evaluated. In the latter, ratings of human performance, robot performance, rushedness, and calmness levels during exercising were evaluated for a rehabilitation robot. Other aforementioned Rating Scale questionnaires have also had applications in pHRC studies [89, 91].

The aforementioned questionnaire employ rating scale methods have served as the primary tool to gauge subjective human experiences. However, with a deeper exploration of understanding human perceptions, it becomes clear that these methods present challenges.

A primary concern is the issue of response bias [93]. Individual rating habits and inclinations introduce variability. There are individuals who naturally assign higher ratings, while others tend toward conservatism, frequently giving lower scores. Additionally, the influence of culture cannot be overlooked. In certain cultures, there is a tendency to avoid extreme scores, whereas in others, participants may gravitate more freely towards the poles of the scale [38]. The discourse around individualism and collectivism within cross-cultural research has been a focal point for understanding the nuances of cultural identities. The construct of individualism-collectivism, as explored through decades of scholarly work, serves as a prism through which cultural variations are examined. Oyserman et al.'s comprehensive meta-analysis brought forth nuanced insights into this domain, suggesting that the construct is not monolithic but rather encompasses multiple facets that could lead to varying magnitudes of cultural differences [94]. Significantly, their analysis challenged the conventional dichotomy that portrays East Asians, particularly Japanese, as collectivists and North Americans as individualists. The discourse on the contrasting tendencies between Japanese participants, who typically shy away from extreme ratings, and American participants, who are more likely to use the entire spectrum of the rating scale, addresses concerns over methodological issues in cross-cultural studies employing Likert scales. These concerns were particularly noted in response to the outcomes of Oyserman et al.'s meta-analysis, which brought to light the potential for methodological biases in survey research that utilizes these scales [94].

Extreme responding refers to a pattern where survey participants consistently choose the most extreme options available, such as consistently rating items with the highest or lowest possible score on a Likert scale, or exclusively selecting "strongly agree" or "strongly disagree" in surveys designed with such response options. This tendency can be influenced by various factors. Cultural background is one significant driver, with evidence suggesting that individuals from certain regions, like the Middle East and Latin America, are more inclined towards extreme responses compared to those from East Asia and Western Europe. Educational background also plays a role, with indications that those with lower levels of educational achievement or intelligence are more susceptible to this bias [95]. Furthermore, the specific phrasing of survey questions can elicit extreme responses, particularly if the questions touch on subjects closely aligned with the respondent's personal beliefs or motivations [95]. Conversely, there is a contrasting bias where respondents preferentially select neutral or moderate options, avoiding any extreme positions. This behaviour represents the opposite end of response bias, where participants gravitate towards the middle of the scale, avoiding more definitive options at either end.

The granularity of rating scales introduces a significant challenge. A high-resolution scale, offering numerous choices, can overwhelm participants as they strive to discern minor differences between options. This overabundance of choices can introduce extraneous variability into the data, as responses may be swayed by factors irrelevant to the study's core aims, thereby increasing bias. Additionally, when the range of options does not fully encapsulate participants' preferences, frustration can ensue due to the lack of accurately representative choices. Conversely, a scale with too few options fails to adequately reflect the subtleties of participants' emotions and experiences. This reduction in scale complexity risks yielding homogeneous outcomes, concealing important variations and intricacies in the data, and thus, could lead to an oversimplified or distorted portrayal of participants' genuine feelings. The simplification in distinguishing between categories poses its own set of issues, including increased difficulty in making selections and the potential trivialization of response categories. This situation may tax participants' cognitive resources, prompt impatience, and amplify the influence of cognitive biases, further compromising the scale's effectiveness and accuracy [96].

Another challenge is also prevalent: the risk of careless or disengaged responses. Participants who are not fond of surveys or find them tedious might rush through the questions, leading to potentially skewed or non-representative results. This haphazard response can reduce the reliability and validity of the data [40, 41].

The RS method could bring inaccurate measurement of subjective impression during pHRC experiment which will affect the evaluation of the control method or pHRC design expecially when the differences in what is being measures are small and nuanced. This motivates work exploring how suitable Likert style methods are in pHRC research, and if alternatives are preferred.

2.5 Concluding Remarks

In this chapter, the fundamental concepts and classifications pertaining to HRI and, more specifically, pHRI are presented. The nature and role of collaborative robots are explored, along with the intricate dynamics of trust within pHRC. Methods of role arbitration and measuring subjective impressions in pHRC contexts are also reviewed. The subsequent chapter will delve into a detailed examination of robot trust in their human co-workers.

Chapter 3

Computational Model of Robot Trust in Human Co-worker

3.1 Introduction

Building on the foundational concepts presented in the previous chapters, this chapter delves deeper into a pivotal aspect of HRI to address research sub-question 1: the computational modelling of robot trust in its human co-workers. While the importance of human trust in robots is widely recognised, there is a compelling case to be made for the reciprocal - the robot's trust in humans. This dimension is especially pertinent in contexts such as pHRC.

Trust is the key to achieving successful HRI. Besides the trust of the human co-worker in the robot, which has been well researched, as detailed in Section 2.2.1, the trust of the robot in its human co-worker should also be considered. A computational model of a robot's trust in its human co-worker for pHRC is proposed. The trust model is a function of the human co-worker's performance which can be characterised by factors including safety, robot singularity, smoothness, physical performance, and cognitive performance. Experiments are conducted with a collaborative robot to verify the trust model developed.

The organisation of this chapter is as follows. The computational model of robot trust in a human co-worker is described in Section 3.2. The human co-worker performance modelling is presented in Section 3.3. An experimental testbed and the design of the experiments are presented in Section 3.4. The results and discussion are shown in Section 3.5 and the conclusion in Section 3.6.

3.2 Computational Trust Model

The trust of humans in robots is dynamic and highly depends on robot performance [75]. Similarly, the dynamics of the trust of robots in humans could mimic the trust of humans in robots. We introduce a real-time computational robot-to-human trust model:

$$T[n] = \frac{\sum_{k=0}^{N} \beta^k p[n-k]}{\sum_{k=0}^{N} \beta^k} \tag{3.1}$$

This model is based on the rationale that more recent performance has a greater impact on the level of trust [5]. In the model, $T \in [0, 1]$ is the trust of the robot in the human co-worker. $T = 0$ represents no trust in the human co-worker, whereas $T = 1$ represents complete trust. $p \in [0, 1]$ is the human co-worker performance calculated in Equation 3.3. A discount factor $\beta \in [0, 1]$ is introduced that represents the effect of past performance $(p[n-1], p[n-2]...p[n-N])$ on trust T. β^k is the weighting of p at the time step $n-k$. When k is larger, β^k is smaller. N is the length of the moving time window and n is the current time step. β and N determine the sensitivity of T to past performance. This model allows past performance to have less effect on T if β and N are small.

3.3 Human Co-worker Performance Modelling

To evaluate human co-worker performance p in the context of pHRC in Equation 3.1, factors that affect human co-worker performance in pHRC need to be identified and quantified. In this research, the human co-worker performance is characterised by the following measures:

- Safety Performance $p_S \in [0, 1]$
- Singularity Performance $p_{SP} \in [0, 1]$
- Smoothness Performance $p_{SM} \in [0, 1]$
- Physical Performance $p_{PW} \in [0, 1]$
- Cognitive Performance $p_{CP} \in [0, 1]$

Detailed explanation on why these factors should be considered is provided in the following sections. All these factors are defined as normalised non-dimensional numbers in the

interval [0,1]. The method proposed by Tran [77] is used to incorporate the different performance factors into a single measure of user performance p:

$$p[n] = \prod_{A=1}^{N_c} p_A^c[n](C + (1-C) \times \sum_{A=1}^{N_{nc}} \gamma_A p_A^{nc}[n]) \tag{3.2}$$

where p_A^c and p_A^{nc} are the critical and non-critical performance factors, respectively. Critical factors p_A^c are strongly relevant to the continuation of a task, the safety of the human co-worker and the robot. Non-critical factors p_A^{nc} are used to assess the performance of human co-worker that is not essential to the task and safety of the human co-worker and the robot, but are still influential when determining trust. N_c and N_{nc} are the number of critical and non-critical performance factors. γ_A is the weighting coefficient and represents the relative importance of each p_A^{nc}, with $\sum_{A=1}^{N_{nc}} \gamma_A = 1$. The parameter C represents the maximum contribution of p_A^c to the overall performance measure.

In this thesis, the Safety Performance p_S and the singularity performance p_{SP} are related to safety. Therefore, p_S and p_{SP} are regarded as p_A^c (critical). The other three performance factors (smoothness performance p_{SM}, physical performance p_{PW}, and cognitive performance p_{CP}) are not related to safety and are regarded p_A^{nc} (non-critical). Consequently, based on Equation 3.2, the human co-worker performance is modelled as:

$$p[n] = p_S[n] \cdot p_{SP}[n](C + (1-C)(\gamma_{SM} \cdot p_{SM}[n] + \gamma_{PW} \cdot p_{PW}[n] + \gamma_{CP} \cdot p_{CP}[n])) \tag{3.3}$$

Weighting coefficients $\gamma_{SM} + \gamma_{PW} + \gamma_{CP} = 1$ are positive constants that could be adjusted based on the relative importance of the corresponding p_A^{nc} according to specific task requirements. Methods for determining these parameters are discussed in later sections.

3.3.1 Safety Performance

In order to achieve safe pHRC, collisions between the robot, human and surrounding objects need to be avoided to prevent potential damage and physical injury. As a result, a safety performance measure p_S is defined based on the possibility of collision between the robot and surrounding objects. The possibility of collision increases from low ($p_S = 1$) to high ($p_S = 0$). Therefore, an interpolation method is employed to acquire a smooth curve of p_S [97]:

$$f(c, x^-, x^+, y^-, y^+) = \begin{cases} y^- & \text{if } c \leq x^- \\ y^+ & \text{if } c \geq x^+ \\ f_p(c, x^-, x^+, y^-, y^+) & \text{otherwise} \end{cases} \tag{3.4}$$

where $f_p(c, x^-, x^+, y^-, y^+)$ is a fifth-order polynomial with null first and second derivatives at x^- and x^+ and the value of f is bounded in the range between y^- and y^+. Based on Equation 3.4, the Safety Performance p_S is calculated as:

$$p_S[n] = f(a[n], a^-, a^+, 1, 0) \tag{3.5}$$

$$a[n] = \frac{v^2[n] - v_0^2[n]}{2\Delta s[n]} \tag{3.6}$$

The value $a[n]$ is the magnitude of the constant deceleration required to stop the robot when it reaches the position to collide with an object which is based on the kinematic equation 3.6. The parameter a^- is the threshold deceleration at which p_S starts to reduce, as shown in Figure 3.1, and the parameter a^+ is the maximum deceleration allowed. The value $v_0[n]$ is the current robot velocity towards the object. The value $v[n]$ is the velocity of the robot when it reaches the position to collide with an object, which is assumed to be zero ($v = 0m/s$) to nullify the impact. $\Delta s[n]$ is the distance between the robot and the object.

Employing a takes into account both the robot velocity toward an object v_0 and the distance between the robot and an object Δs. When Δs is small or v_0 is large, a is large, indicating that the possibility of collision is high. These characteristics make a a useful measure to encapsulate the performance of the human user with respect to safety.

Figure 3.1 shows the use of Equation 3.4 to smoothly calculate Safety Performance with deceleration a. The measure p_S starts to reduce once $a > a^-$ as the possibility of collision increases. The measure p_S reaches the minimum value ($p_S = 0$) once $a \geq a^+$. p_S is bounded in the range between 0 ($y^+ = 0$) and 1 ($y^- = 1$).

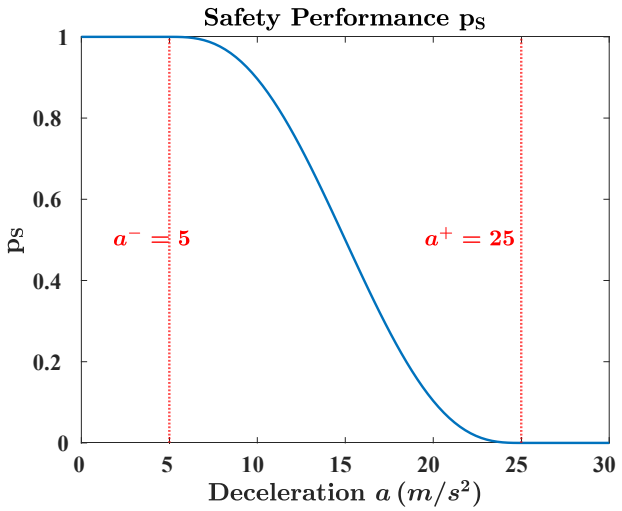


FIGURE 3.1: Example of the safety performance p_S used to normalize the value of deceleration a which is shown in Equation 3.4. $x^- = a^- = 0$, $x^+ = a^+ = 25$, $y^- = 1$, $y^+ = 0$

The safety performance measure p_S needs to be evaluated for each surrounding object i using Algorithm 1, where N_{obj} is the total number of surrounding objects.

Algorithm 1 Safety Performance Calculation

for $i \leftarrow 1$ **to** N_{obj} **do**
  $\vec{d_i} = \vec{s_i} - \vec{s}_{robot}$ {Vector from position of robot $\vec{s}_{robot}$ to position of object i $\vec{s_i}$}
  $\vec{v_i} = 0$ {The velocity toward object i}
  {If the angle between vector of robot velocity $\vec{v}_{robot}$ and d_i is less than or equal to 90 degree, calculate the velocity toward the object i}
  if $\vec{v}_{robot} \cdot \vec{d_i} \geq 0$ **then**
    $\vec{v_i} = proj_{\vec{d_i}} \vec{v}_{robot}$
  end if
  $a_i = \frac{0^2 - \|\vec{v_i}\|^2}{2 \times \|\vec{d_i}\|}$ {Deceleration for object i based on Equation 3.6.}
  $p_S^i = f(|a_i|, a^-, a^+, 1, 0)$ {Safety Performance for object i based on Equation 3.5.}
end for
$p_S = \min(p_S^i), \forall i \in 1, 2, 3...N_{obj}$ {Safety Performance p_S is defined as the smallest safety performance among all the surrounding objects}

3.3.2 Singularity Performance

Figure 3.2 demonstrates a example of the manipulator entering a singular configuration when a human co-worker holds the robot through a handlebar and moves around to complete a task. It is assumed that when this form of pHRC is applied to performing tasks,

for example, in a workplace setting, the human co-worker's focus is primarily on how to get the task done quickly with expected quality. With their focus being on moving the end effector to perform the task, and not on the robot itself, the human is likely to move the robot to its maximum reach without notice. Therefore, the robot manipulator is likely to move to singular configurations.

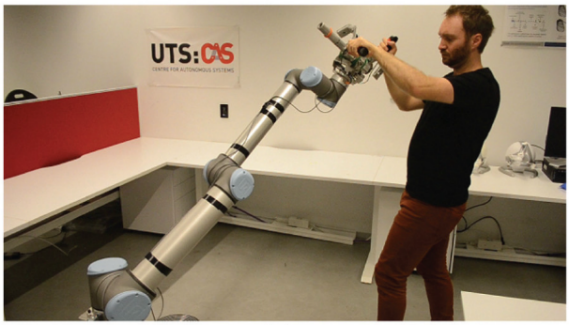


FIGURE 3.2: Example of the manipulator entering a singular configuration [2].

When the manipulator is close to a singular configuration, small motions of the end-effector can cause the joint velocities to be large, causing the robot behaviour to be unpredictable and dangerous. As a result, monitoring whether the manipulator configuration is approaching singularity is a key to maintaining the safety of human co-workers.

It is reasonable to assume that users with enough experience can operate the robot through much of the workspace while keeping the robot away from singularity, even with their attention being primarily on the task at the end effector. Based on this, Singularity performance $p_{SP} \in [0, 1]$ is proposed to quantify the human performance in avoiding singular configurations. When $p_{SP} = 1$ (or $p_{SP} = 0$), the possibility of problems occurring due to proximity to a singular configuration is extremely low (or high).

The performance measure p_{SP} is defined as a function of the smallest singular value σ_{min}. σ_{min} can be calculated by singular values decomposition (SVD) of the Jacobian matrix (J). Inspired by [2], an exponentially shaped function was introduced to scale the singular value.

$$p_{SP}[n] = \begin{cases} 1 - \varphi^{\frac{\sigma_{min}[n]-\sigma^-}{\sigma^+-\sigma^-}} & \sigma_{min}[n] > \sigma^- \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

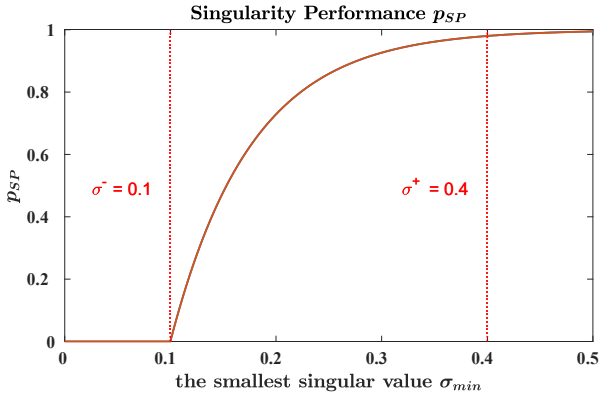


FIGURE 3.3: Example of singularity performance p_{SP} versus the smallest singular value σ_{min} which is shown in Equation 3.7. $\sigma^+ = 0.25$, $\sigma^- = 0.1$ and $\gamma = 0.02$.

The parameter σ^+ is the smallest singular value threshold at which p_{SP} starts to reduce as shown in Figure 3.3. Alternatively, σ^- is the minimum value allowed for the smallest singular value ($\sigma^+ > \sigma^- > 0$). The difference between σ^+ and σ^- should not be too small to ensure a smooth transition of p_{SP} with respect to σ_{min}. φ determines the smoothness of the transition between σ^- and σ^+ ($1 >> \varphi > 0$). $\varphi = 0.02$ should be tuned to ensure a smooth transition and $p_{SP}(\sigma^+) \approx 1$.

From Figure 3.3, when $\sigma_{min} > \sigma^+$, $p_{SP} \approx 1$, which indicates that the possibility of problems due to the proximity to a singular configuration is low. When $\sigma_{min} < \sigma^-$, $p_{SP} = 0$, which indicates that the possibility of problems due to proximity to a singular configuration is high.

In addition, whether the manipulator is approaching or leaving a singular configuration is also another important factor to be considered when calculating the Singularity Performance p_{SP}. p_{SP} should be higher when the manipulator is heading away from a singular configuration because the possibility of entering a singular configuration is lower compared with heading toward a singular configuration. Based on [2], if the desired velocity of the end effector of the robot in Cartesian space x_d causes a decrease in the smallest singular value σ_{min}, it indicates that the manipulator is approaching a singular configuration and vice versa. Therefore, the approximate joint positions $q[n+1]$ at the next time step are defined as:

$$q[n+1] = q[n] + \delta J^{-1} x_d, \text{ for small } \delta \tag{3.8}$$

δ is the length of time looking ahead into the future to linearly extrapolate the future configuration of the manipulator. Readers are directed toward [2] to acquire a more detailed explanation of Equation 3.8.

The Jacobian matrix $J[n+1]$ in the next time step is calculated based on $q[n+1]$. $\sigma_{min}[n+1]$ is calculated on the basis of $J[n+1]$ through SVD. Therefore, $\sigma^+ = \sigma^+_{AW}$ and $\sigma^- = \sigma^-_{AW}$ when the manipulator is leaving a singular configuration ($\sigma_{min}[n+1] \geq \sigma_{min}[n]$) and $\sigma^+ = \sigma^+_{AP}$ and $\sigma^- = \sigma^-_{AP}$ when approaching a singular configuration ($\sigma_{min}[n+1] < \sigma_{min}[n]$):

$$\sigma^+, \sigma^- = \begin{cases} \sigma^+_{AW}, \sigma^-_{AW} & \text{if } \sigma_{min}[n+1] \geq \sigma_{min}[n] \\ \sigma^+_{AP}, \sigma^-_{AP} & \text{if } \sigma_{min}[n+1] < \sigma_{min}[n] \end{cases} \tag{3.9}$$

The parameters in Equation 3.9 should be tuned such that $\sigma^+_{AP} > \sigma^+_{AW} > 0$ and $\sigma^-_{AP} > \sigma^-_{AW} > 0$ to ensure $p_{SP}(\sigma_{min}|\sigma^+_{AW}, \sigma^-_{AW}) > p_{SP}(\sigma_{min}|\sigma^+_{AP}, \sigma^-_{AP})$. This is to reflect that heading towards a singular configuration is considered an indicator of lower human performance, and vice versa.

3.3.3 Smoothness Performance

Balasubramanian [98] proposed that smoothness can be used to assess the control ability of a human co-worker in pHRC. An experienced human co-worker normally operates the robot with smoother movements compared to novice users [98]. Consequently, the smoothness of robot motion is one of the factors for evaluating the human co-worker performance during pHRC. One way to measure smoothness is to use jerk, the first derivative of acceleration [99]. There is evidence to suggest that human motion involves the minimisation of the jerk. Therefore, jerk is an effective measurement of smoothness [100]. As a result, jerkiness jk can be calculated by [101]:

$$jk[n] = \|\ddot{\vec{v}}[n]\| \tag{3.10}$$

where $\vec{v}[n]$ is the vector of the linear velocity of the end effector of the robot in Cartesian space.

Smoothness performance $p_{SM} \in [0, 1]$ is used to quantify the degree of smoothness when the human co-worker moves the robot during performing a pHRC task. $p_{SM} = 0$ is defined as an extremely unsmooth movement. $p_{SM} = 1$ corresponds to when the smoothness of human co-worker's movement is within an acceptable range. The interpolation function in Equation 3.4 is used to generate p_{SM}:

$$p_{SM}[n] = f(jk[n], jk^-, jk^+, 1, 0) \tag{3.11}$$

where jk^- is the threshold jerkiness at which p_{SM} begins to reduce from 1, and jk^+ is the maximum smoothness allowed, as shown in Figure 3.4.

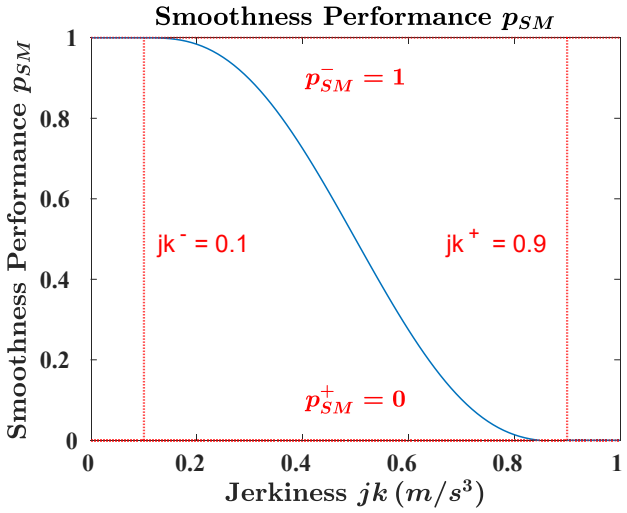


FIGURE 3.4: Example of the Smoothness Performance p_{SM} used to normalize the value of jerkiness jk which is shown in Equation 3.11. $x^- = jk^- = 0.1$, $x^+ = jk^+ = 0.9$, $p_{SM}^- = 1$, $p_{SM}^+ = 0$

3.3.4 Physical Performance

The human co-worker's performance is also affected by the physical workload the human co-worker has to take during a period of time. The high workload causes muscle fatigue and therefore affects performance. Human physical performance is an important element in pHRC as a result of continuous force exchange between the human and the robot. Sadrfaridpour [102] proposed a model of human physical performance for collaborative manufacturing. In pHRC, the interaction between humans and robots is mostly based on force exchange. Hence, Sadrfaridpour's model can be used for pHRC because the force applied by the human co-worker could be measured through a force-torque sensor in real time.

Based on the model of [102], the human physical performance, p_{PW}, is calculated by Equation 3.12:

$$p_{PW}[n] = \frac{F_{max,iso}[n] - F_{th}}{MVC - F_{th}} \tag{3.12}$$

where F_{th} is the equilibrium point at which fatigue and recovery balance out. Maximum Voluntary Contraction (MVC) is the maximum isometric force with zero fatigue level. Assume that the co-worker starts without fatigue, then $F_{max,iso}[0] = MVC$. $F_{max,iso}[n]$ is

the maximum isometric force which will reduce when a human co-worker's muscle applies a force for some time due to the level of fatigue increasing. Therefore, the physical performance p_{PW} degrades correspondingly when $F_{max,iso}$ decreases, as shown in Equation 3.12. $F_{max,iso}$ is calculated as [102]:

$$\begin{aligned} F_{max,iso}[n+1] = F_{max,iso}[n] - C_f F_{max,iso}[n]\frac{\|F[n]\|}{MVC} \\ +C_r(MVC - F_{max,iso}[n]) \end{aligned} \tag{3.13}$$

where C_f and C_r are the fatigue and recovery coefficients which are individual-dependent. $\|F[n]\|$ is the magnitude of force applied by the human co-worker. F_{th} is defined as:

$$F_{th} = MVC\frac{C_r}{2C_f}(-1+\sqrt{1+\frac{4C_f}{C_r}}) \tag{3.14}$$

The calculated maximum isometric force $F_{max,iso}$, and human physical performance measure p_{PW}, are shown in Figure 3.5. It is noted that human physical performance is extremely complex, and the physical performance model used in this work is a simplified representation. More complex strength or fatigue models can refer to [33].

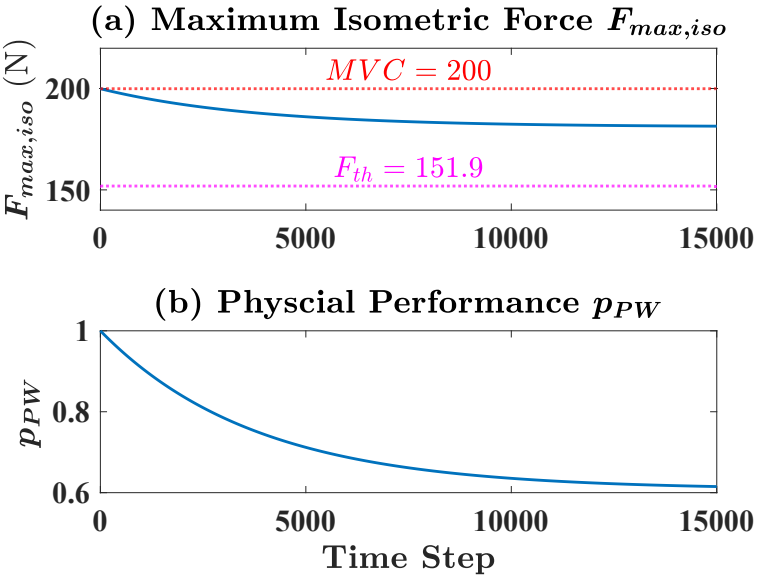


FIGURE 3.5: (a) Example of the maximum isometric force $F_{max,iso}$ versus time step n (Equation 3.13). (b) Example of physical performance p_{PW} versus time step n (Equation 3.12). $MVC = 200$, $C_f = 10^{-4}$, $C_r = 2.4 \times 10^{-4}$ and $F_{th} = 151.9$ with constant applied force $\|F[n]\| = 50N$.

3.3.5 Cognitive Performance

Cognitive workload is a measure of mental work when performing a task. Therefore, it is believed to have an effect on human performance [103]. Sadrfaridpour [102] proposed using Yerkes-Dodson (YD) law to describe the cognitive performance of the human co-worker. YD law [104] stated that when the level of arousal increases, the human cognitive performance increases accordingly. However, this increase in cognitive performance will only occur before the level of arousal increases up to the point known as the optimal level of arousal (OLA). After this point, human cognitive performance reduces as the level of arousal increases. Bertrand [105] proposed that task difficulty affects the relationship between arousal and cognitive performance. The OLA value is smaller for a more difficult task. Saeidi [106] proposed a dynamic model, which describes the relationship between human arousal, task difficulty, and performance based on the YD law.

$$p_{CP}[n] = (p_{CP}^{+} - p_{CP}^{-})(\frac{r[n]}{\beta_{CP}})^{\beta_{CP}}(\frac{1-r[n]}{1-\beta_{CP}})^{1-\beta_{CP}} + p_{CP}^{-} \tag{3.15}$$

where $r[n] \in [0,1]$ is the utilisation ratio of the human co-worker, which represents the amount of time that the human co-worker has been controlling the robot. p_{CP}^{+} and p_{CP}^{-} are the upper and lower limits of cognitive performance p_{CP} that are individual-specific. The value of β_{CP} is determined by the difficulty of the task. A larger β_{CP} represents a less difficult task. The most difficult task is when $\beta_{CP} = 0$. Figure 3.6 shows the human performance compared to different task difficulty β_{CP}. When the utilisation ratio $r[n]$ gradually increases, p_{CP} increases to the highest point (OLA point), then decreases. In addition, the OLA point shifts to the right when the task becomes easier ($\beta_{CP} \to 1$).

$$\begin{aligned} &r[n+1] = a_r r[n] + b_r M[n] \\ &a_r = 1 - \frac{1}{\tau} \in (0,1) \quad b_r = \frac{1}{\tau} \in (0,1) \end{aligned} \tag{3.16}$$

where τ is the time constant that defines the sensitivity of the next time step utilisation ratio $r[n+1]$ to the current utilisation ratio $r[n]$. The larger τ causes less variation in the utilisation ratio r. $M[n]$ is the control mode, where $M[n] = 1$ refers to the manual control mode, and $M[n] = 0$ represents the autonomous control mode [106]. In manual control mode, the human co-worker's utilisation of the robot continuously increases. Alternatively, in autonomous control mode, the human co-worker's utilisation of the robot continuously decreases.

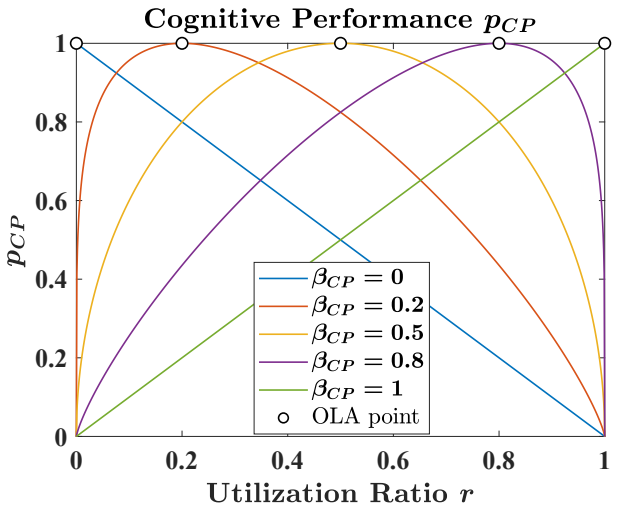


FIGURE 3.6: Examples of the Cognitive Performance p_{CP} versus utilization ratio r with different levels of task difficulty β_{CP} (Equation 3.15).

3.4 Experiments

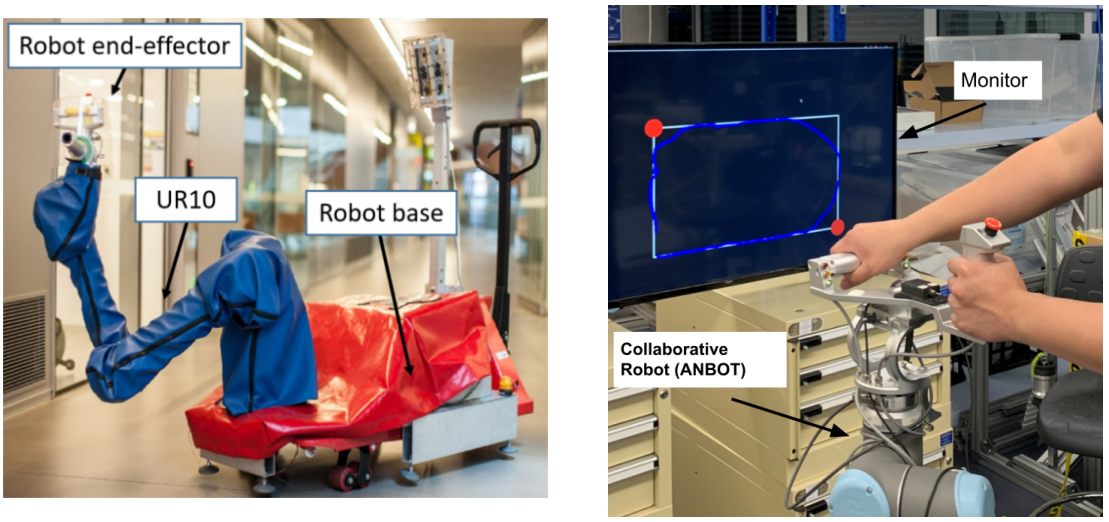
This section details the design and results of three experiments to verify the proposed robot-to-human trust model by using a human-robot collaborative system. In these experiments, the smoothness, singularity, safety, physical, and cognitive performance measures are individually evaluated.

3.4.1 Experimental Testbed

The experiment testbed is called ANBOT [1] (Figure 3.7a) which is a collaborative robotic system for the collaborative operation of humans and robots. Besides a sensor package for perception and operational and environmental awareness, it also consists of a sophisticated software package for user intention recognition, safe physical interaction and control. ANBOT consists of a UR10 arm from Universal Robots and a six-axis force-torque sensor mounted between the robot end-effector and the robotic arm to measure the interaction forces applied by the human co-worker.

The human co-worker constantly holds the handlebar mounted on the robot end effector, as shown in Figure 3.7b. In front of the human and robot is a monitor on to which the direction of the end-effector is projected, shown graphically as a dot. Also shown on the monitor is a trajectory that the human co-worker is expected to follow by controlling the motions of the robot. The movement of the robot is restricted to two dimensions to reduce

the complexity of the experiment. Hence, the human co-worker can only move the robot in vertical and horizontal directions, which are parallel to the monitor.

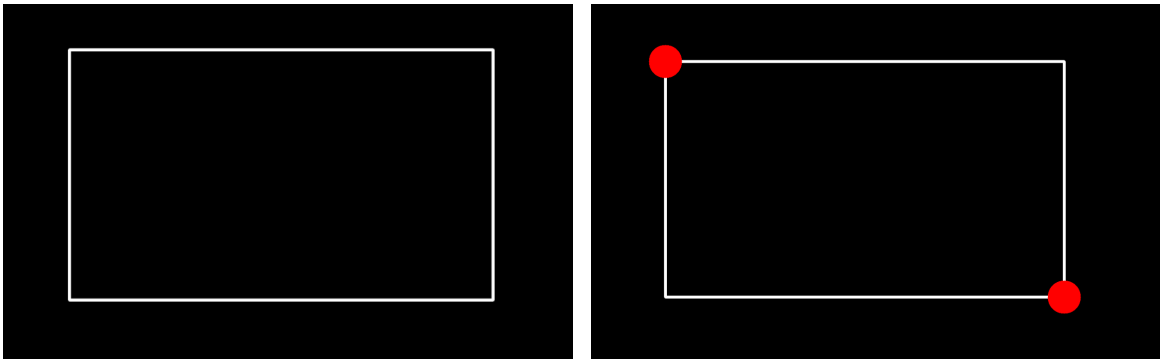


(a) (b)

FIGURE 3.7: (a) Experimental Testbed - ANBOT. (b) The human co-worker is operating the ANBOT to follow a desired trajectory.

3.4.2 Design of Experiments

Three experiments are designed for verifying the proposed computational trust model. Figure 3.8a shows the trajectory that the robot end-effector needs to track in Experiments 1 and 2. Figure 3.8b shows the trajectory for Experiment 3. The big filled red circles represent objects (or obstacles) located on the trajectory. The white line is the desired trajectory that needs to be followed. Each of the experiments starts and ends at the bottom left corner of the rectangle with clockwise movement.



(a) (b)

FIGURE 3.8: (a) Trajectory tracking path for Experiments 1 and 2. (b) Trajectory tracking path for Experiment 3.

Experiment 1: This experiment is designed to verify the smoothness performance model in Section 3.3.3 and the Singularity Performance model in Section 3.3.2. In this experiment, the human subject is asked to move the robot end effector to follow the path (shown in Figure 3.8a) as accurate and smooth as possible to avoid a significant change of velocity. There is no limitation in time to complete the entire path. During the path following, the robot manipulator will approach and leave singular configurations. Therefore, the Singularity Performance model can be verified.

Experiment 2: This experiment is to verify the physical performance model in Section 3.3.4 and the cognitive performance model in Section 3.3.5. Physical performance p_{PW} and cognitive performance p_{CP} are mainly affected by the duration of operation of the robot. p_{PW} is also affected by the magnitude of the interaction force. Therefore, the human subject is asked to move along the path (Figure 3.8a), which is similar to Experiment 1, but as quickly as possible to ensure that the human subject exerts a large force during the execution of the task. Additionally, the human subject is asked to continuously move around the path for five loops to ensure that the duration of Experiment 2 is long enough to cause fatigue.

Experiment 3: This experiment is designed to verify the safety performance model in Section 3.3.1 by introducing two objects/obstacles on the same path in Experiments 1 and 2. In addition to avoiding objects, the requirement for the human subject is the same as in experiment 1. However, the human subject is required to move as fast as possible from the top right corner of the path towards the top left corner to demonstrate the effect of the velocity component of Equation 3.6 on Safety Performance p_S. This experiment is also conducted to verify the combined models of human co-worker performance in Section 3.3 and the computational trust model in Section 3.2 because the variation of all performance factors can be observed.

This experiment follows the procedure approved by the UTS Ethics Committee with approval number ETH21-6346. Results from a single subject are analysed since these experiments aim to verify that the models work as designed. Validation in terms of their use for human-robot role arbitration is performed in subsequent chapters.

3.5 Results and Discussion

3.5.1 Experiment 1 – verifying the Smoothness Performance model and the Singularity Performance model

3.5.1.1 Singularity Performance

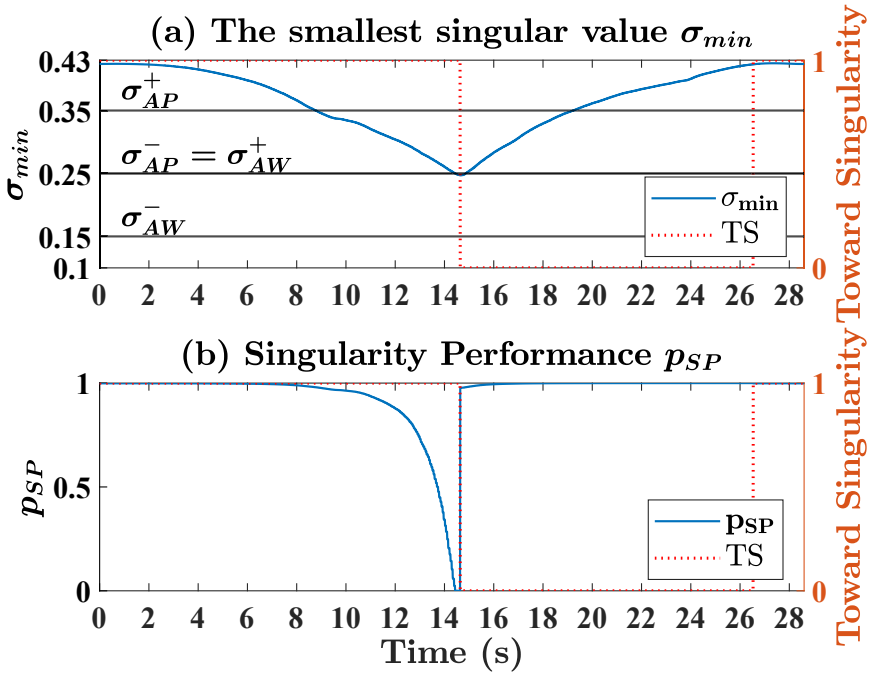


FIGURE 3.9: (a) The smallest singular value σ_{min}. (b) Singularity performance p_{SP}. TS is the prediction on whether the robot is heading towards a singular configuration based on Equations 3.8 and 3.9. $TS = 1$ (or $TS = 0$) refers to heading towards (or Away) a singular configuration. Note that $TS = 1$ when $\sigma_{min}[n+1] < \sigma_{min}[n]$ and $TS = 0$ when $\sigma_{min}[n+1] \geq \sigma_{min}[n]$.

The values of the parameters in the Singularity Performance model are $\sigma^+_{AW} = 0.25$, $\sigma^-_{AW} = 0.15$, $\sigma^+_{AP} = 0.35$, $\sigma^-_{AP} = 0.25$, [107]. $\varphi = 0.02$ and $\delta = 0.1$ are suggested [2].

Figure 3.9 shows the smallest singular value σ_{min} and the singularity performance p_{SP} in Experiment 1. Before 8.5s, σ_{min} is larger than $\sigma^+_{AP} = 0.35$, hence $p_{SP} \approx 1$ (Equation

3.7). At around 8.5s, p_{SP} starts to decrease because $\sigma_{min} < \sigma^{+}_{AP}$ until $p_{SP} = 0$ due to $\sigma_{min} < \sigma^{-}_{AP}$ at around 14.5s. After this time, the robot starts to head away from a singular configuration ($TS = 1 \rightarrow 0$) which can be seen that σ_{min} starts to increase, then p_{SP} increases up to 1 instantaneously due to $\sigma_{min} > \sigma^{+}_{AW}$. Equations 3.8 and 3.9 successfully predict whether the manipulator is heading toward a singular configuration.

3.5.1.2 Smoothness Performance

The smoothness jk in Equation 3.10 is calculated by fitting the second time derivative of the second degree polynomial curve of the velocity using the least-squares method with a time window of 1 second for noise reduction.

Figure 3.11 shows that jk is much higher at the beginning of the experiment and at the turning points (corners) of the desired path (Figure 3.10). The human co-worker needs to decelerate before a turning point and accelerate after passing the turning point, which causes the unsmooth movement. jk is high at the beginning of the experiment because the human co-worker starts to move from rest. In order to show the details of Figure 3.11, jk is bounded between $[0, 1]m/s^3$.

The lower and upper thresholds of jerkiness jk^{-} and jk^{+} depend on the smoothness requirement. jk^{-} and jk^{+} should be set to small when the requirement is high. In this experiment, the requirement for smoothness of human co-worker's movement is high. Therefore, $jk^{-} = 0.1m/s^3$ and $jk^{+} = 0.9m/s^3$ (Equation 3.11). In Figure 3.11, the Smoothness Performance p_{SM} decreases correspondingly when jk increases as long as $jk > jk^{-}$. It can also be seen that $p_{SM} = 0$ when $jk >= jk^{+}$ at the beginning of the experiment.

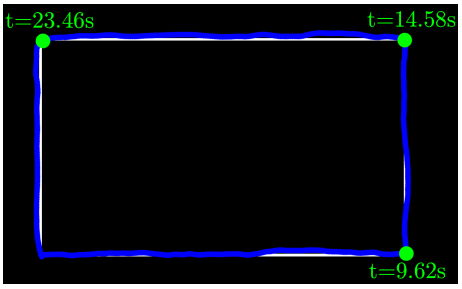


FIGURE 3.10: The blue line corresponds to the actual robot end-effector Trajectory. The labelled time corresponds to the time when the human co-worker changes the directions on the corners of the trajectory.

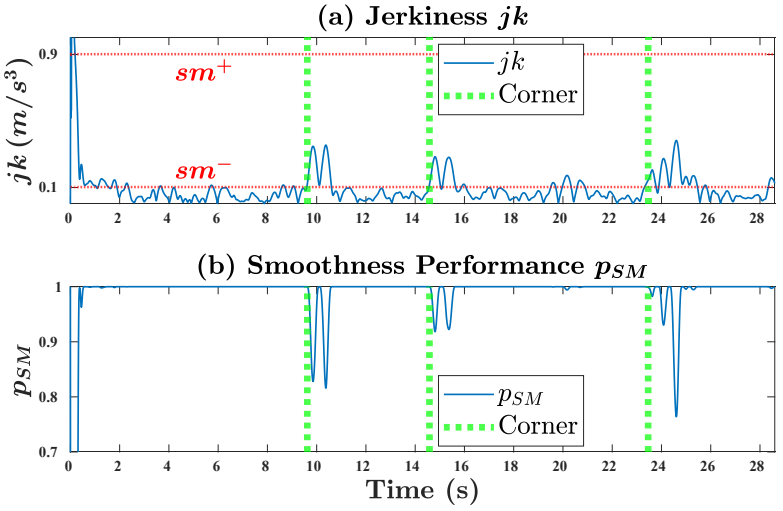


FIGURE 3.11: (a) Jerkiness jk.(b) Smoothness Performance p_{SM}. The green vertical lines correspond to the time when human co-worker changes the directions at the corners of the trajectory.

3.5.2 Experiment 2 – verifying the Physical Performance model and the Cognitive Performance model

3.5.2.1 Physical Performance

The values of parameters in the physical performance model are set $MVC = 200$, $F_{th} = 151.9$, $C_f = 10^{-4}$ and $C_r = 2.4 \times 10^{-4}$ which are similar to the literature [102]. However, the parameters are individual-dependent. The detailed method for measuring those parameters can be found in [108].

Figure 3.12 shows that the magnitude of interaction force $\|F\|$, the maximum isometric force $F_{max,iso}$ and physical performance p_{PW} versus time in Experiments 1, 2 and 3. In Experiment 2, the human subject applied much larger force and executed the task for the longest time (around 40.6s) because the human subject is required to move along the trajectory as fast as possible for five loops. It can be seen that $F_{max,iso} = 196.16N$ and $p_{PW} = 0.92$ are the lowest at the end of experiments which represent the highest level of fatigue. The results demonstrate that the Physical Performance model captures the behavour of reducing physical performance over time due to fatique from prolonged interaction with the robot.

Figure 3.12 also shows that the value of maximum isometric force $F_{max,iso}$ starts to reduce from maximum voluntary contraction ($MVC = 200$) at the beginning of the experiments. At the end of the experiments, $F_{max,iso} = 198.97N, 196.16N, 198.88N$ in Experiments 1,

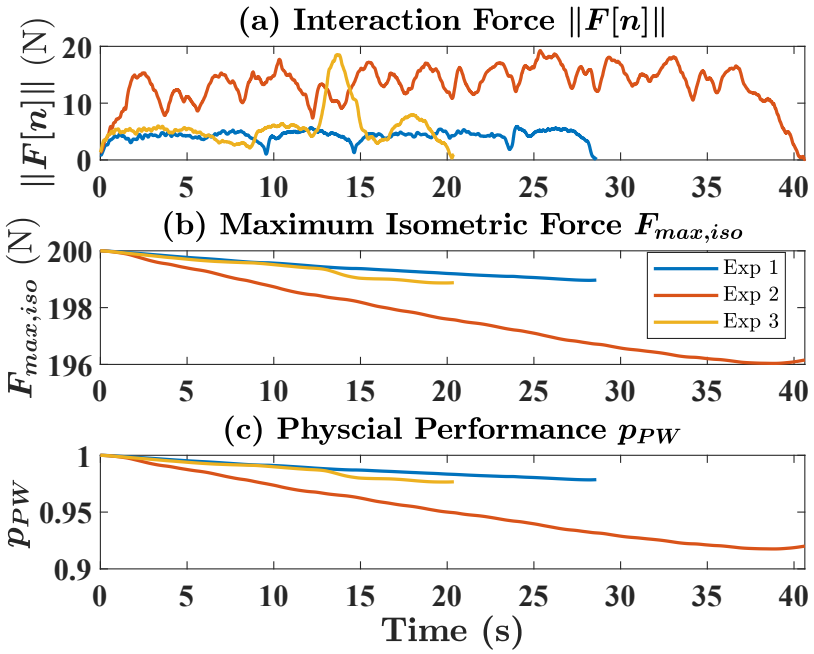


FIGURE 3.12: (a) Magnitude of interaction force $\|F\|$ applied by human co-worker and measured by force-torque sensor. (b) The maximum isometric force $F_{max,iso}$. (c) Physical Performance p_{PW}.

2 and 3 and those values are close to 200 because the duration of all the experiments are short (28.6s, 40.6s, 20.4s), which indicates that the fatigue level of human co-worker increases slightly.

3.5.2.2 Cognitive Performance

The parameters in the cognitive performance model are subject-dependent and task-dependent. Details on how to select the parameters can be found in [106]. Based on [106], the parameter values used in this experiment are $p_{CP}^{-} = 0.391$, $p_{CP}^{+} = 0.4602$, $\beta_{CP} = 0.74$ and $a_r = 0.9991$. The parameter a_r depends on the system sampling rate, which in [106] is 10Hz. The sampling rate for this experiment is 125Hz, which corresponds to the value $a_r = 0.999928$. Because the human subject constantly moves the robot during the experiment, the control mode is always 1, i.e. $M = 1$ in Equation 3.16.

As shown in Figure 3.13, the utilisation ratio r starts to increase from 0 when the human subject starts to operate the robot. r constantly increases during the experiment because r represents the amount of time that the human co-worker has operated the robot. At the end of the experiments, $r \approx 0.3$ in Experiment 2 is the highest because the time the human subject has controlled the robot is the longest (around 40.6 seconds). Cognitive

performance p_{CP} starts to increase correspondingly from the lower threshold of cognitive performance $p_{CP}^{-} = 0.391$. The reason why p_{CP} continuously increases during the experiment is that r is lower than the optimal level of arousal (OLA), which is when $\beta_{CP} = 0.74$. The reason why r and p_{CP} are the same between $0s$ and $20.4s$ for the three experiments is that all cognitive performance parameters are the same $(a_r, p_{CP}^{-}, p_{CP}^{+}, \beta_{CP})$ and the human subject constantly operates the robot ($M = 1$ in Equation 3.16).

It is noted that the results obtained in Figure 3.13 are similar to the result in [106].

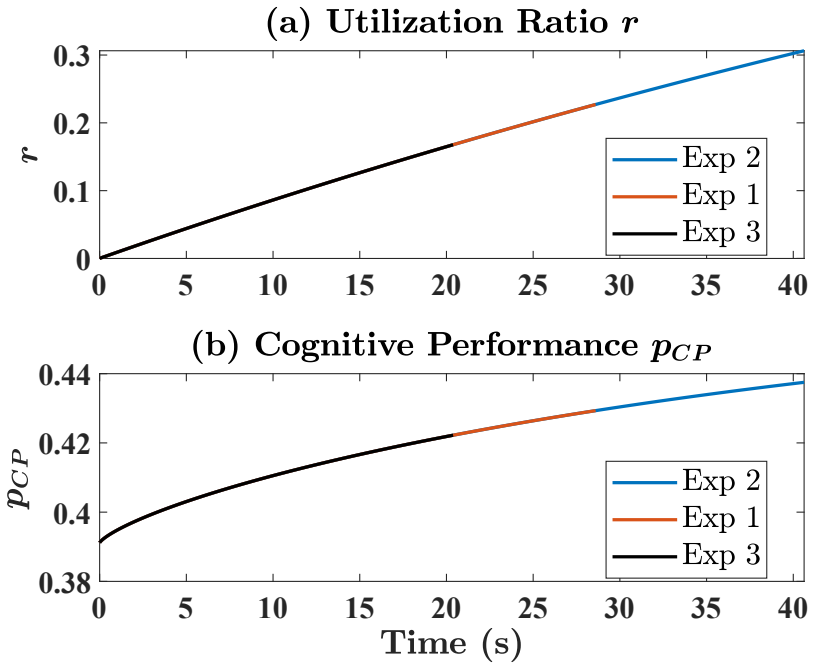


FIGURE 3.13: (a) Utilization ratio r versus time and (b) Cognitive performance p_{CP} in Experiments 1, 2 and 3.

The effect of task difficulty β_{CP} on p_{CP} was previously shown in Figure 3.6. The effect of M on r and p_{CP} has not been investigated because the human subject constantly operates the robot during the experiment ($M = 1$). Hence, a simulation is conducted with the same parameters of the cognitive performance model used in this experiment by switching the control mode between manual $M = 1$ and autonomous $M = 0$ during an experiment. The results are shown in Figure 3.14. At 60s, $M = 1 \rightarrow 0$, r decreases indicating the amount of time the human co-worker operates the robot decreases. Consequently, p_{CP} decreases accordingly. At 180 s, $M = 0 \rightarrow 1$, r, and p_{CP} start to increase again.

Another simulation was run that shows the effects of a longer interaction between the human and the robot. This is to demonstrate the situation where over-utilisation of the robot, that is, r is greater than the OLA point, leads to decreased cognitive performance, as shown in Figure 3.6.

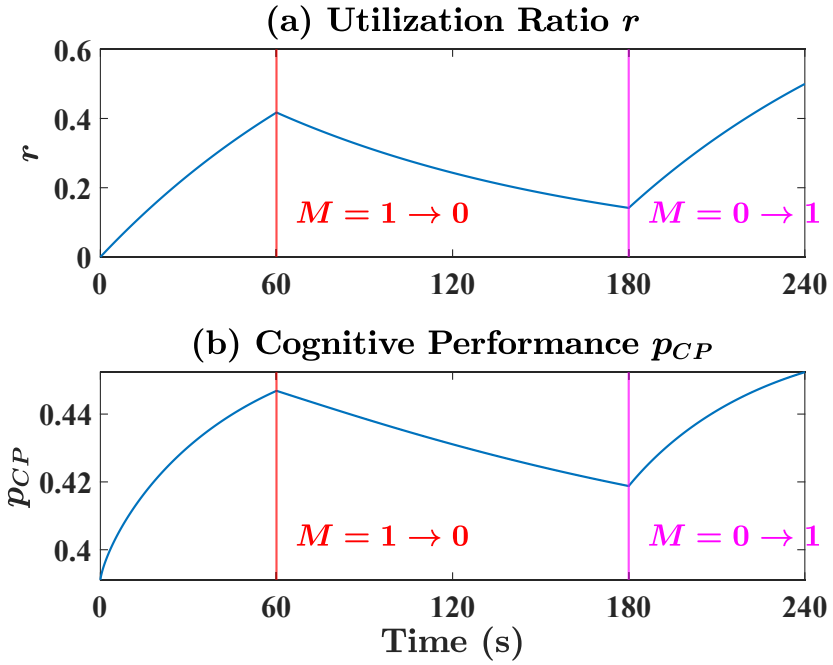


FIGURE 3.14: Simulation results: (a) Utilization ratio r and (b) Cognitive Performance p_{CP} with varying control model M when $p_{CP}^{-} = 0.391$, $p_{CP}^{+} = 0.4602$, $\beta_{CP} = 0.74$ and $a_r = 0.999928$. $M = 1$ refers to manual control mode and $M = 0$ refers to autonomous control mode.

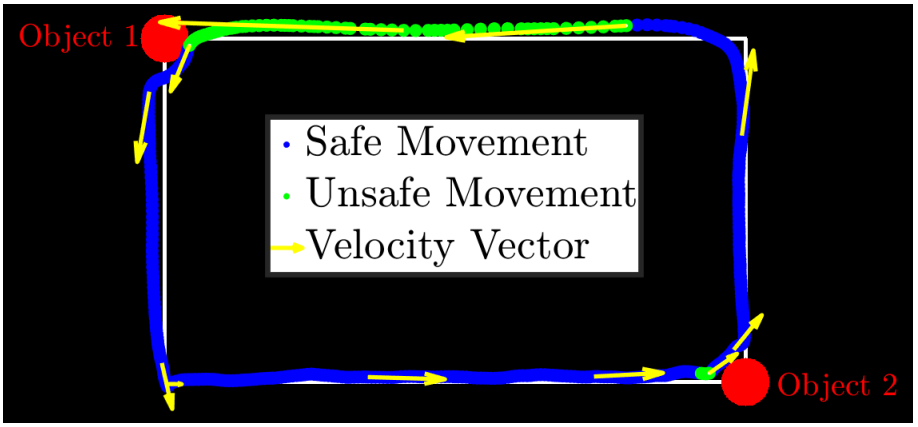
3.5.3 Experiment 3 – verifying the Safety Performance model, Human Co-worker Performance and the Computational Trust Model

3.5.3.1 Safety Performance

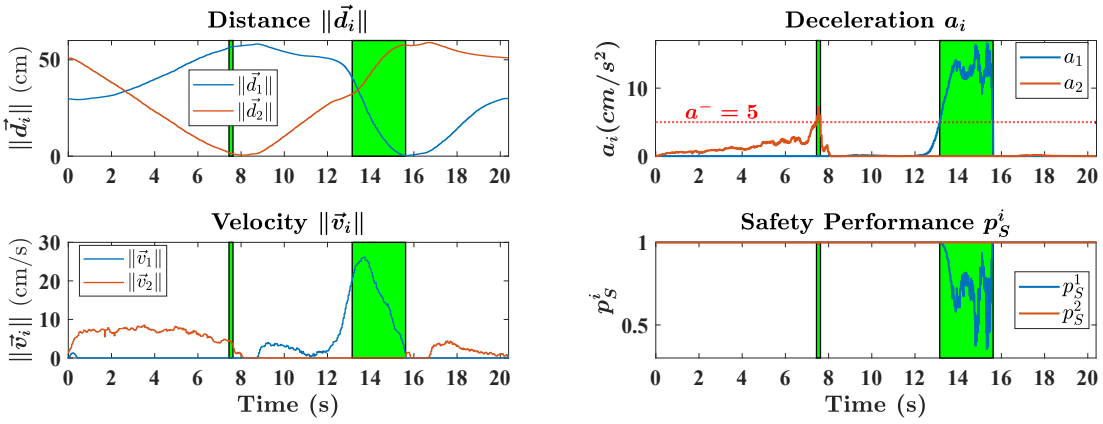
In Figure 3.15a, the green section indicates the period in which the movement of a human co-worker is regarded as unsafe. The blue section is regarded as a safe movement. The period of unsafe movement for approaching Object 2 (0.18s) is much shorter than that for Object 1 (2.47 s) because the velocity toward Object 1 is much larger than that of Object 2, as shown in Figure 3.15b. It can also be seen in Figure 3.15a that the length of the robot velocity vector is much longer in the green section when approaching Object 1.

Movement is identified as unsafe only when the robot is very close to Object 2 (1.15cm). Therefore, when the velocity is low, the distance to the objects will be the main factor determining the safety of movement.

Another essential factor for determining the safety movement is the direction of motion of the robot. At around 15.62s, the robot is heading away from Object 1 as shown by the yellow arrow in Figure 3.15a. Movement is regarded safe even though the distance between the robot and Object 1 is small (0.22cm) as shown in Figure 3.15b.



(a) The actual robot end-effector trajectory along the designed trajectory



(b) (c)

FIGURE 3.15: (a) The actual robot end-effector trajectory along the designed trajectory. The blue trajectory refers to the safe movement ($\max(a_1, a_2) \leq a^-$) and the green trajectory refers to the unsafe movement ($\max(a_1, a_2) > a^-$) which corresponds to the shaded green regions in Figures 3.15b and 3.15c. The yellow vectors are the robot velocity vectors. (b) Top: Distance between the robot and Object 1 $\|\vec{d}_1\|$ (Object 2 $\|\vec{d}_2\|$). Bottom: the velocity of the robot towards Object 1 $\|\vec{v}_1\|$ (Object 2 $\|\vec{v}_2\|$). (c) Top: The magnitude of the constant deceleration required to stop the robot when it reaches the position to collide with Object 1 a_1 (Object 2 a_2). Bottom: The safety performance for Object 1 p_S^1 (Object 2 p_S^2).

The parameter values used are $a^- = 5cm/s^2$ and $a^+ = 25cm/s^2$ (Equation 3.5). These parameters can be tuned to smaller values when the requirement on safety is higher. Smaller values are recommended to increase the robustness of Safety Performance measurement when there are measurement errors of positions of surrounding Objects introduced by the sensors. In Figure 3.15c, the safety performance p_S^i decreases correspondingly when $a_i > a^-$, indicating that the possibility of a collision starts to increase. It is hard to observe the decrease in p_S^2 because a_2 is a little over a^-.

3.5.3.2 The combined Human Co-worker's Performance

The values of γ_{SM}, γ_{PW} and γ_{CP} in Equation 4.6 are determined by the relative importance of the non-critical performance factors p_A^{nc}. For this experiment, the requirement on the smoothness of movement is high. In addition, the duration of the experiment is short, resulting in the variation of p_{PW} and p_{CP} being subtle. Hence, we set $\gamma_{SM} = 0.8$, $\gamma_{PW} = 0.1$ and $\gamma_{CP} = 0.1$. p_A^{nc} is important due to the smoothness requirement of human co-worker is high, therefore, $C = 0$.

Figure 3.16 shows that the human performance (p) is low at the beginning of the experiment because of poor smoothness performance ($p_{SM} = 0$), and smoothness is highly weighted ($\gamma_{SM} = 0.8$). It is also seen that p decreases immediately when p_S (Figure 3.15c) or p_{SP} reduces, because they are both critical performance factors (Equation 4.6). The performance is $p < 1$ at all time due to p_{CP} having bounds within the interval $[0.391, 0.4602]$ during the experiment, as shown in Figure 3.13. The effect of p_{PW} on p is minimal because the variation of p_{PW} is small (Figure 3.12) and $\gamma_{PW} = 0.1$.

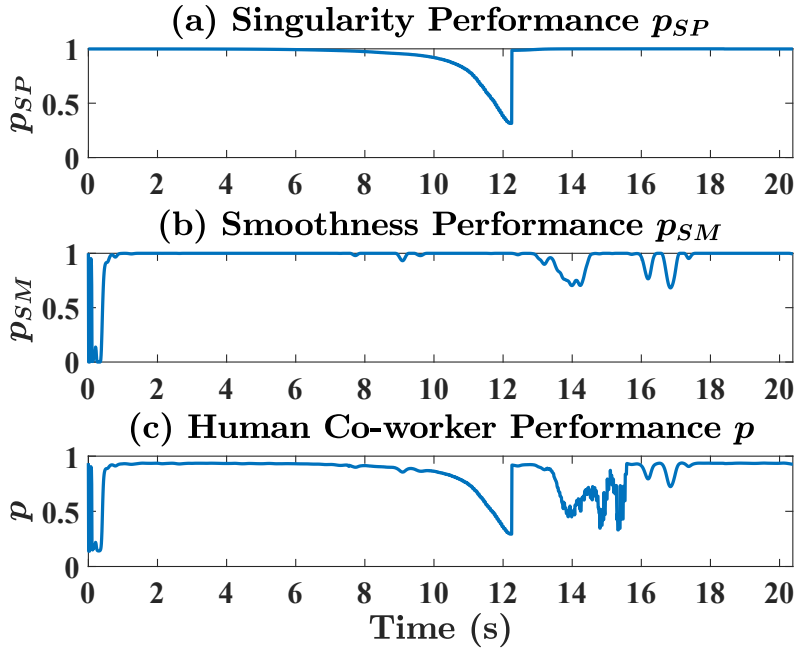


FIGURE 3.16: (a) Singularity Performance p_{SP}. (b) Smoothness Performance p_{SM}. (b) The combined Human co-worker performance p

3.5.3.3 The Computational Trust Model

The robot-to-human trust T is based on the measure of human performance, with parameter β weighing the importance of recent performance levels during a window of N

performance measures (Equation 3.1). The trust with different values of β and N is simulated as shown in Figure 3.17 based on the combined human co-worker performance (Figure 3.16). The time interval of p is bounded into [10 14]s to show the comparisons more clearly.

A pattern can be observed that smaller values of β and N will result in robot-to-human trust T being more responsive to the variation of p. This provides system developers with the ability to tune the responsiveness of the trust model based on the needs and risks of the pHRC task, with β and N being smaller when the risk level is higher. Readers are directed to Table 5.1 in [77] to determine the risk level of the corresponding pHRC application. Moreover, larger values of β and N will result in T becoming smoother. A good choice of values for β and N should balance the smoothness and responsiveness of T to p.

If the past performance is not important for evaluating T, β can be designed small so that $T \approx p$. In this case, the value of N will not affect T. If $\beta^k \approx 0$, N should be adjusted to ensure $N < k$ because it will be computationally expensive to calculate the historical performance terms with weightings close to zero $\beta^k \approx 0$.

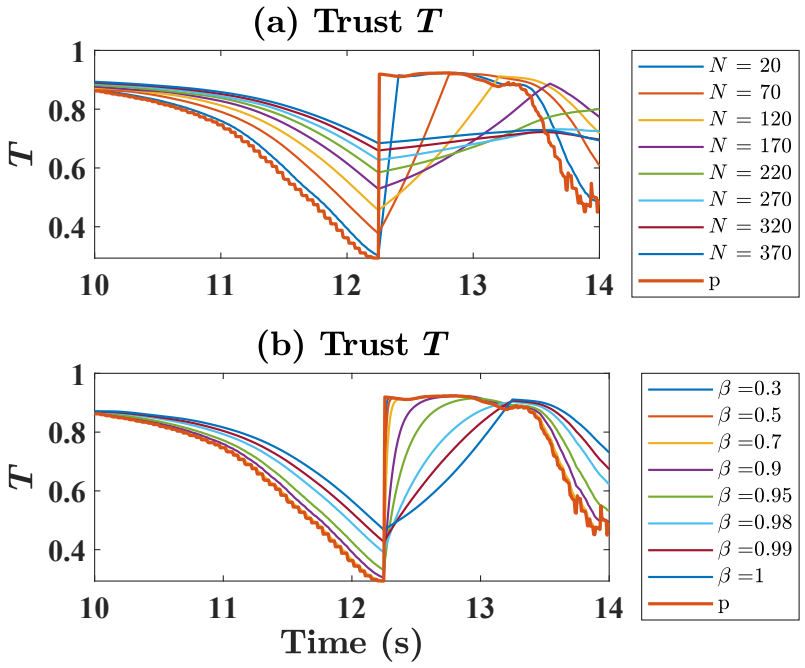


FIGURE 3.17: (a) Trust plots for different values of time window size N when $\beta = 0.999$. (b) Trust plots for different values of β when $N = 125$.

3.6 Conclusion

A computational model of robot trust in a human co-worker was proposed. This model takes into account many factors in physical human-robot collaboration, including robot safety, robot singularity, smoothness of robot motion, and human physical and cognitive performance. Three experiments were conducted to verify all the factors in the model. The rationale behind the different performance measures was presented, and the desired behaviour was described. Each performance measure was then individually verified, confirming that the performance models produced the expected output during a real human-robot physical task. Furthermore, the performance measures were combined into a single trust model.

Although each component of the robot-to-human trust model was verified, fitting the model to the human co-worker, or using the resulting trust value to influence the human-robot interaction was not yet performed. However, trust, though crucial, represents only one dimension of this intricate relationship. As the discussion progresses, it becomes clear that trust alone does not govern the dynamics of HRI. An equally significant facet is robot self-confidence – an internal assessment of a robot's perception of its own capabilities. Alongside this, there is a necessity to highlight how the intertwined concepts of trust and self-confidence guide the essential process of role arbitration in pHRC. This alignment extends beyond understanding individual parameters; it is about discerning their collective influence on dynamic interactions during real-time collaboration. In transitioning into the next chapters of the thesis, the focus will shift to this intriguing intersection.

Chapter 4

Role Arbitration using Robot Trust and Self-Confidence

4.1 Introduction

Building on the foundational understanding of a robot's trust in its human co-worker as established in the previous chapter, this chapter delves deeper into the dynamic interplay between robot trust, self-confidence, and their collective influence on role arbitration within pHRC. As has been discerned, trust alone does not dictate the equilibrium of control; instead, a robot's intrinsic self-confidence plays a pivotal role in these decision-making processes.

In scenarios where trust in the human is paramount but coupled with a robot's diminished self-confidence, the collaborative dynamics naturally skew towards human dominance. In contrast, when a robot exhibits heightened self-assurance or confidence, it assumes a more assertive stance, taking the reins of control. Balance, or perhaps tug of war, between trust and self-confidence is crucial to shaping the future of pHRC as shown in Figure 4.1.

Role arbitration, therefore, extends beyond the mere allocation of responsibilities; it is a constantly evolving relationship influenced by trust in the partner, trust in one's self (i.e. self-confidence), and real-time evaluations of performance and capability. This chapter seeks to unravel this complex interaction, leading to a solution for achieving a harmonious, efficient, and fluid human-robot partnership.

In this chapter, sub-research questions Q2 and Q3 are being addressed. Role arbitration in pHRC refers to how control is allocated between humans and robots for a task and

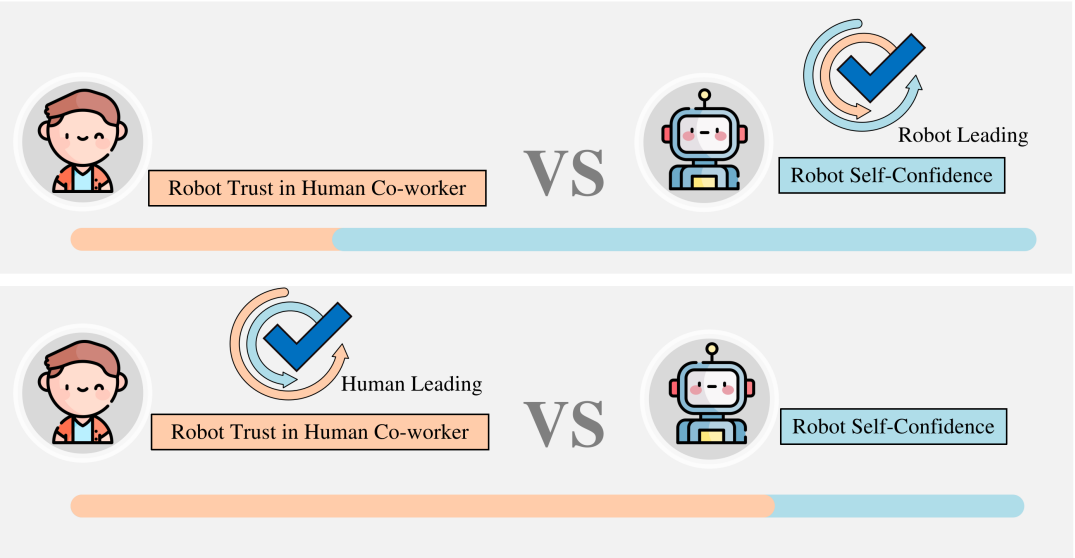


FIGURE 4.1: A robot trust and self-confidence role arbitration method for physical collaboration with human co-workers. This approach hinges on the principle that when a robot's trust in its human co-worker exceeds its own self-confidence, the human partner assumes greater control of the task. Conversely, if the robot's self-confidence surpasses its trust in the human, then the robot takes on a more dominant role in the control.

allows combining the strengths of humans and robots to improve combined performance and reduce human physical or cognitive effort [30].

The willingness of a human to rely on the robot taking control during pHRC depends on the difference between human trust in the robot and the human's trust in themselves to perform the task [34]. The robot can take control when human trust in the robot is higher than trusting themselves; otherwise, the human would remain in control. Saeidi [109] proposed a human trust in robots and self-confidence model and a control switching method for (semi)autonomous mobile robotic systems. Saeidi [110] also proposed a mixed-initiative method based on human trust in robots.

The human co-worker interaction experience with robots is one of the critical aspects of pHRC due to the constant physical contact between the human and the robot [111]. The aforementioned trust-based role allocation strategies are based on human-to-robot trust [109, 110]. In Chapter 3 a computational model for robot trust in human co-workers in pHRC was proposed and verified. To emulate the human control allocation process, we want the robot to act as an active partner, i.e. a peer. Therefore, in this chapter we develop a human-preferred role arbitration method based on TSC.

The proposed TSC model is defined as the difference between robot-to-human trust and the robot's self-confidence. In this thesis, the robot-to-human trust model, developed in Chapter 3 and published in [112], is used. This trust model takes into account safety, robot

singularity, smoothness, physical performance, and cognitive performance. The robot's self-confidence is modelled based on whether a human co-worker agrees with the control actions of the robot. The robot's self-confidence is high when a human agrees with robot control, and vice versa. In addition, when the robot's trust in a human co-worker is much higher than the robot's self-confidence, the control is allocated to the human co-worker and vice versa. However, the way the control is allocated when the robot's trust in human co-workers and itself are both high or low becomes a question. In this case, the history of robot self-confidence can be used. If the robot's self-confidence in history is high, it indicates that the robot control was satisfied by the human in the past. Hence, the control will be biased toward the robot and vice versa.

Therefore, a TSC-based role arbitration method is proposed and evaluated with the aim of showing through experimental analysis that this method can achieve superior combined human-robot performance, reduce the workload of human co-workers and improve subjective preference. The control system for role arbitration method is presented in Section 4.2. The computational TSC model is described in Section 4.3. The TSC-based role arbitration method is shown in Section 4.4. An experimental testbed and the design of the experiments are presented in Section 4.5. The results and discussion are shown in Section 4.6, and the conclusions are given in Section 4.7.

4.2 Control System

Evrard [113] proposed a system that could continuously switch between two different behaviours (follower and leader) for the robot in pHRC. Based on this concept, a control scheme is proposed that consists of both follower and leader roles of the robot as shown in Equation 4.1. The roles (follower and leader) of the robot is adapted by a role arbitration parameter $\alpha \in [0, 1]$ which will be explained in Section 4.4. Humans act in a leader role (that is, in control) and robots as followers when $\alpha = 0$. Inversely, the robot acts as a leader and the human as a follower when $\alpha = 1$.

$$\dot{x} = \dot{x}_H + \alpha \dot{x}_R \tag{4.1}$$

The term $\dot{x}$ is the velocity command sent to the robot manipulator. $\dot{x}_H$ is the human control and $\dot{x}_R$ is the robot control. It is noted that when $\alpha = 0$ only the human control signal is enacted; however, when $\alpha = 1$ both the human and robot control signals are combined. This means that the human control signal is always considered when computing the velocity command, which was a deliberate design decision based on human safety.

The follower role of the robot is achieved through an admittance control such that the movements of the robot end-effector are based on the force, F_H, applied by the human co-worker when the human is in control (the leader, $\alpha = 0$) which is shown as:

$$M_d\ddot{x}_H + D_d\dot{x}_H = F_H \tag{4.2}$$

M_d and D_d are the desired robot inertia and damping matrix. F_H is the external force applied to the robot end effector by a human co-worker. When the robot is leader ($\alpha = 1$), a bounded impedance control model inspired by [2] is employed:

$$\dot{x}_R = min(\dot{x}_{max}, \dot{x}_{max}\frac{d}{d_{th}}) \tag{4.3}$$

d is the distance between the desired task-dependent position and the current position of the end effector. $\dot{x}_{max}$ is the maximum robot velocity. d_{th} is the threshold distance when the robot control $\dot{x}_R$ reaches $\dot{x}_{max}$, i.e. $d = d_{th}$. The bounded $\dot{x}_R$ helps to improve the safety of human co-workers and interaction experiences during pHRC due to the physical couplings of the human-robot system [2]. Hence, $\dot{x}_{max}$ should be designed to ensure safety and an acceptable interaction experience. In addition, $\dot{x}_{max}$ should be large enough to ensure that the robot acts as a leader when $\alpha = 1$.

4.3 Robot Trust in human co-worker and Robot Self-Confidence (TSC) Model

The willingness of a human to rely on the robot taking control depends on the difference between human-to-robot trust and human self-confidence to perform the task themselves [34]. Hence, a similar control allocation process, but from the perspective of the robot, is proposed. The robot's reliance on a human co-worker taking control depends on the difference between robot-to-human trust and the robot's self-confidence. The robot-to-human trust reflects the capability, from the robot's perspective, of human co-workers to take control during pHRC. The self-confidence of the robot reflects the ability of the robot to take control. Therefore, a real-time computational model of the difference between robot-to-human trust $T \in [0, 1]$ and robot self-confidence $SC \in [0, 1]$ at time step n can be simply defined as:

$$TSC[n] = T[n] - SC[n] \tag{4.4}$$

where $TSC \in [-1, 1]$. T and SC will be introduced in Sections 4.3.1 and 4.3.2, respectively.

4.3.1 Robot Trust in Human Co-worker (T)

In the previous chapter a model for the robot's trust in the human co-worker was developed. This model is summarised here for convenience. The trust is computed as:

$$T[n] = \frac{\sum_{k=0}^{N_1} \beta_1^k p_H[n-k]}{\sum_{k=0}^{N_1} \beta_1^k} \tag{4.5}$$

where $T \in [0, 1]$ is the robot's trust in a human co-worker. $T = 0$ (or $T = 1$) represents no trust (or full trust) in the human co-worker. $p_H \in [0, 1]$ is a normalised measure of human performance. A discount factor $\beta_1 \in [0, 1]$ was introduced to reduce the sensitivity to historical performance, as the most recent performance has a greater impact on trust [5]. β_1^k is the weighting of p_H at the time step $n-k$. N_1 is the length of the moving time window and n is the current time step. p_H is modelled by important pHRC factors, including safety p_S, singularity p_{SP}, smoothness p_{SM}, physical p_{PW} and cognitive performance p_{CW} [112].

$$p_H[n] = p_S p_{SP}(C + (1 - C)(\gamma_{SM} p_{SM} + \gamma_{PW} p_{PW} + \gamma_{CP} p_{CP})) \tag{4.6}$$

The weighting coefficients $\gamma_{SM} + \gamma_{PW} + \gamma_{CP} = 1$ are positive constants that could be adjusted based on the relative importance of the corresponding factors according to the specific requirements of the task. $1 - C$ represents the maximum contribution of p_{SM}, p_{PW}, and p_{CW}.

4.3.2 Robot Self-Confidence (SC)

Robot self-confidence $SC \in [0, 1]$ is defined as the robot's trust in itself. Similar to Equation 4.5, the robot self-confidence is defined as:

$$SC[n] = \frac{\sum_{k=0}^{N_2} \beta_2^k p_R[n-k]}{\sum_{k=0}^{N_2} \beta_2^k} \tag{4.7}$$

β_2 and N_2 are similar to β_1 and N_1 which have been explained in Section 4.3.1.

The rationale behind using human-robot agreement as a measure of self-confidence is based on the idea that assuming the human is probably in a better position than the robot to make the ultimate decision about who should lead due to its superior cognitive advantage.

The robot performance p_R is defined as the level of the agreement of the human co-worker on the control action of the robot $\dot{x}_R$. There are scenarios, such as a human co-worker changing a task plan, accumulation of sensor noise, or unexpected obstacles that will result in the robot not accomplishing the pre-programmed task by itself or does not behave as the human co-worker expected. This will result in the level of agreement on $\dot{x}_R$ being low, which leads to p_R being low and vice versa.

p_R is modelled as a function of the difference between human control $\dot{x}_H$ and robot control $\dot{x}_R$ as shown in Equation 4.8. When the difference between those two controls is too significant, and the two control directions are opposite, it indicates that the level of agreement with $\dot{x}_R$ is low. If those two controls are in the same direction, it indicates that the human co-worker agrees with the $\dot{x}_R$. Therefore, the difference between those two controls, V_D, is shown as:

$$V_D[n] = \begin{cases} |\dot{x}_H - \dot{x}_R| & \text{if } sign(\dot{x}_H) \neq sign(\dot{x}_R) \\ 0 & \text{Otherwise} \end{cases} \tag{4.8}$$

When the human and robot are in agreement, $V_D[n]$ is small, and when in disagreement, $V_D[n]$ is large. The value of V_D at any given time step may not represent the actual human co-worker intention to agree or disagree with the robot because human co-workers may unintentionally apply a force in the opposite direction or a large force due to uncertainty of human motor movement. Therefore, the moving averaged method is employed as shown in Equation 4.9. Hence, the averaged difference M_{V_D} is calculated by:

$$M_{V_D}[n] = \frac{1}{W_r} \sum_{i=n-w_r+1}^{n} V_D[i] \tag{4.9}$$

W_r is the time window that determines the sensitivity of M_{V_D} on V_D. If W_r is small, M_{V_D} may be sensitive to the random movement of a human co-worker. If W_r is large, a human co-worker needs to apply additional force (Equation 4.2) to signal the disagreement. Robot performance p_R is then defined as:

$$p_R = \begin{cases} 1 & M_{V_D} \leq M^-_{V_D} \\ \frac{M^+_{V_D} - M_{V_D}}{M^+_{V_D} - M^-_{V_D}} & M^-_{V_D} < M_{V_D} < M^+_{V_D} \\ 0 & M_{V_D} \geq M^+_{V_D} \end{cases} \tag{4.10}$$

$M^+_{V_D}$ and $M^-_{V_D}$ are the two thresholds of M_{V_D}. When $M_{V_D} \leq M^-_{V_D}$, M_{V_D} is small indicating the human co-worker agrees with $\dot{x}_R$ so that p_R is high ($p_R = 1$) and vice versa. When the values of $M^+_{V_D}$ and $M^-_{V_D}$ are large, the human co-worker will be required to put more physical effort in order to take control of the robot. The advantage is that the robot

control will not be easily affected by accidental human movement. Furthermore, when those values are small, the control is easily shifted toward the human co-worker. However, robot control can be easily affected by unintentional human movement.

4.4 A Robot Trust and self-confidence based Role Arbitration Method (α)

Saedi [110] proposed a TSC-based control allocation method for mobile robot. However, a step-like arbitration function was proposed, which is not appropriate in pHRC because human comfort will be significantly affected when the control authority is shifted suddenly [85].

A novel smooth role arbitration function $\alpha(TSC)$ is proposed. A fifth-order polynomial equation (Figure 4.2) is proposed with null first and second derivatives at the TSC^-, and TSC^+, which is inspired by [97]. Hence, the role arbitration function $\alpha(TSC)$ is defined as a function of TSC:

$$d\alpha(TSC) = \begin{cases} 1 & TSC \leq TSC^- \\ 0 & TSC \geq TSC^+ \\ aTSC^5 + bTSC^4 + cTSC^3 + dTSC^2 + eTSC + f & \text{Otherwise} \end{cases} \tag{4.11}$$

$\alpha(TSC)$ is a fifth-order polynomial with null first and second derivatives at TSC^- and TSC^+. TSC^- and TSC^+ are the thresholds determining whether the robot is acting as a leader ($\alpha = 1$ when $TSC <= TSC^-$) or follower ($\alpha = 0$ when $TSC >= TSC^+$). When the robot trust in human is much larger than trust in itself ($T >> SC$), TSC is large ($TSC >= TSC^+$ based on Equation 4.4). The robot's reliance on the human is high and as a result is acts as a follower ($\alpha = 0$), and vice versa.

A dilemma arises when robot trust in human co-worker and robot trust in itself are both high or low (When both T and SC approach 1 or 0) resulting in the difference between them approaches 0 ($TSC \rightarrow 0$) from Equation 4.4. The role allocation has not been addressed yet by the scientific community.

Hence, a method is proposed to address this problem. As can be seen from the example shown in Figure 4.2, with $TSC^- = -0.5$ and $TSC^+ = 0.5$ defined, the arbitration $\alpha(TSC = 0) = 0.5$. If TSC^- and TSC^+ are made smaller, the control is more biased toward human co-workers (α is smaller) across the interval $TSC \in [-1, 1]$.

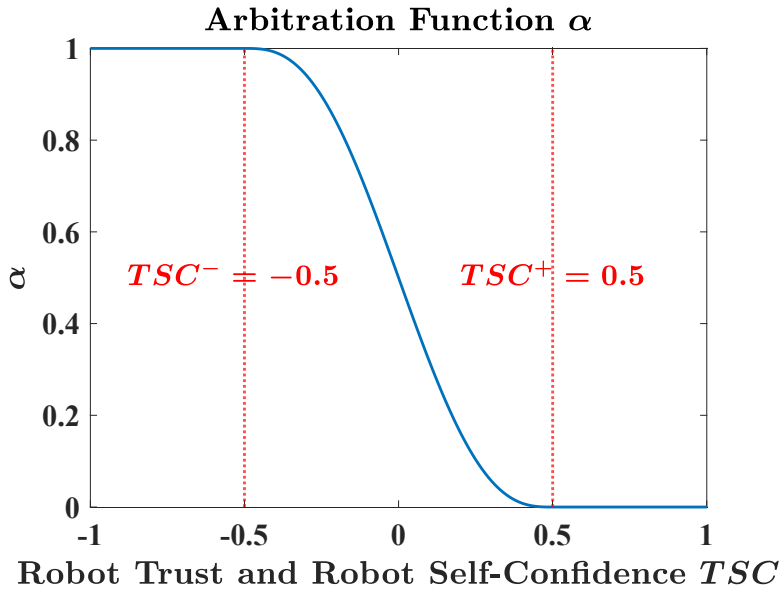


FIGURE 4.2: Examples of the arbitration function α (Equation 4.11) $TSC^- = -0.5$, $TSC^+ = 0.5$, $a = -6$, $b = 0$, $c = 5$, $d = 0$, $e = -1.875$ and $f = 0.5$.

Therefore, instead of estimating a fixed value of TSC^- and TSC^+ heuristically, dynamic equations (Equation 4.12) are developed for adapting $TSC^+ \in [0, 1]$ and $TSC^- \in [-1, 0]$ based on the history of human agreement with the robot control $\dot{x}_R$ to address different levels of robot's capability in accomplishing the task by itself. If $\dot{x}_R$ consistently mismatch human intention (SC is low from Equation 4.10), this represents that the robot cannot accomplish the task by itself. As a result, TSC^+ and TSC^- will converge to 0 and -1, $\alpha \rightarrow 0$ when TSC is close to 0 to ensure the human leader role. If $\dot{x}_R$ consistently matches human intention (SC is high), the robot should take the leader role to maximize the combined performance and reduce human co-worker workload due to the robot's inherent advantage in speed, accuracy, and power. Correspondingly, TSC^+ and TSC^- will gradually increase and converge to 1 and 0 so that $\alpha \rightarrow 1$ when $TSC \approx 0$.

$$\begin{cases} \dot{TSC^+} = \frac{1}{\tau}(SC - TSC^+) \\ \dot{TSC^-} = \frac{1}{\tau}(SC - TSC^-) \\ TSC^- = TSC^- - 1 \end{cases} \tag{4.12}$$

τ is the time constant which determines the sensitivity of $\dot{TSC^+}$ and $\dot{TSC^-}$ to SC. When τ is larger, the variation of TSC^+ and TSC^- will be smaller and vice versa.

Based on the trials on the actual robot, it was found that control authority shifts from robot to human ($\alpha : 1 \rightarrow 0$) instantaneously which is desirable and intuitive for a human co-worker when the human and robot disagree. If the control authority is shifted too slowly,

the human co-worker will likely feel that the robot is uncontrollable. However, smooth and continuous control shifting is desirable from the human co-worker to the robot ($\alpha : 0 \rightarrow 1$). Abrupt control shifting will significantly affect the human's perception [85]. Hence, α is constrained as:

$$\alpha[n] = \begin{cases} \alpha[n-1] + \frac{1}{f_S \times \tau_r} & \alpha[n] - \alpha[n-1] > \frac{1}{f_S \times \tau_r} \\ \alpha[n] & \text{otherwise} \end{cases} \tag{4.13}$$

f_S is the sampling rate of the system. τ_r is the minimum time taken for role arbitration between $\alpha = 1$ and $\alpha = 0$ which is individual-dependent. τ_r is suggested to be greater than 0.48 seconds because it is the fastest human response time [114].

4.5 Experimental Evaluation

4.5.1 Experiment Testbed

The experiment testbed is the same as in Section 3.4.1.

4.5.2 Design of Experiments

A moving target (Red filled circle) tracking experiment is designed for verifying the proposed TSC-based role arbitration method as shown in Figure 4.3. In this experiment, a human co-worker is asked to move the robot end-effector to track the moving target and at the same time avoid collisions with the obstacles (Green filled circles) that are placed on the trajectory of the moving target. The experiment is considered a failure when the robot collides with obstacles. The moving target's position is known, but the obstacle's position is unknown to the robot. The purpose of this experiment design is to replicate a scenario where the human co-worker has superior perceptive and decision making capabilities.

Considering safety, any movement of the end-effector that positions it outside the desired range of the moving target is regarded as unsafe. Therefore, in this experiment, p_S is defined as the possibility of tracking with the desired moving target. As a result, when the possibility of tracking is high, p_S is high. $a^- = 5cm/s^2, a^+ = 10cm/s^2$ based on the expert trials and specific parameter selection method is similar as Section 6.3.1. Smoothness p_{SM}, physical p_{PW} and cognitive performance p_{CW} are not considered important in this experiment design, therefore, $C = 1$ in Equation 4.6. Also, the robot movement will always need to be in non-singular configurations during the experiment, so that $p_{SP} = 1$.

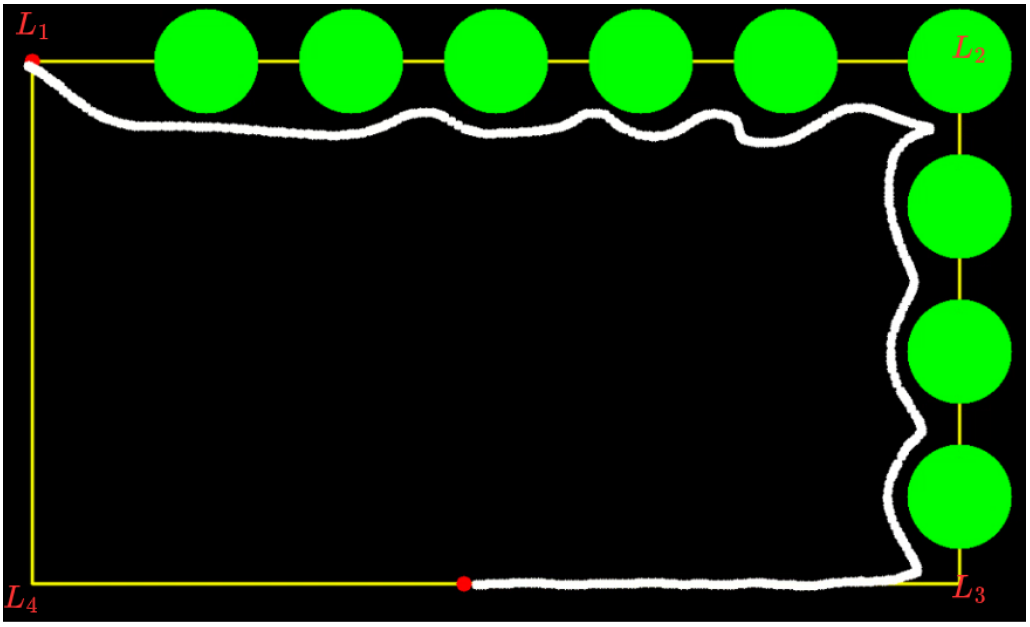


FIGURE 4.3: The target (red-filled circle) starts from the top left corner ($L1$) and moves around the rectangle path (yellow line) in a clockwise direction with constant speed ($L1 \rightarrow L2 \rightarrow L3 \rightarrow L4$). When the moving target returns to the starting position ($L1$), it is regarded as completing one experiment. The big filled green circles are the obstacles. The white line is the actual trajectory of the robot end-effector. L1, L2, L3 and L4 are the timestep when the moving target reaches four corners of the rectangle path.

4.5.3 Three Control Strategies and Evaluations

For comparison of results, three control strategies are implemented in the experiments:

1. *HL - Human Leading*: Human in control $\alpha = 0$ (Equation 4.1)
2. *RL - Robot Leading*: Robot in control $\alpha = 1$ (Equation 4.1)
3. *RA - Role Arbitration*: Robot Trust and Self-Confidence Based Role Arbitration Method $\alpha(TSC)$ (Equation 4.11)

To evaluate the effectiveness of the proposed method, the following metrics are used:

1. *Average Magnitude of Human Force Over the Whole Trajectory*: $\frac{1}{Nc}\sum_{0}^{Nc}(\frac{1}{t_N - t_1}\sum_{t_1}^{t_N}||F_H(t)||)$. Nc is the total number of experiments for the corresponding control strategies. t_1 is the initial timestep of the experiment. t_N is the final timestep. $||F_H(t)||$ is the magnitude of human force at each timestep.
2. *Average Tracking Error Over the Whole Trajectory*: $\frac{1}{Nc}\sum_{0}^{Nc}(\frac{1}{t_N - t_1}\sum_{t_1}^{t_N}||e(t)||)$. $||e(t)||$ is the tracking error which is the distance between the position of the end-effector and the moving target at each timestep.

3. *Average Magnitude of Human Force at each timestep*: $\frac{1}{Nc}\sum_0^{Nc}||F_H(t)||$.
4. *Average Tracking Error at each timestep*: $\frac{1}{Nc}\sum_0^{Nc}||e(t)||$.
5. *Failure Rate*: $FR = \frac{\text{number of Failed experiments}}{\text{Total number of experiments}}$. Failure experiments refer to the situation when the robot collides with obstacles.
6. *NASA TLX*: Post-experiment questionnaires for measuring the perceived workload of the participants.

4.5.4 Participants and Testing Procedure

Fifteen subjects participated in the experiment. Ten participants are from the UTS Robotics Institute and five participants are from the general public. The ages of the participants range between 20 and 40. After completing the consent form, each participant learns how to operate the collaborative robot system as shown in Figure 3.7a until they are familiar with the task. This is to reduce the learning effect. After these steps, each participant will undertake five trials for each control strategy in a random order to reduce the order effect. The participants are required to answer the questionnaires once they finish the experiments. This experiment follows the procedure approved by the UTS Ethical Committee with approval number ETH21-6346.

4.6 Results and Discussion

Figure 4.4 shows the mean and standard deviation of the human co-worker's average tracking error and magnitude of force (Metrics (1) and (2) in Section 6.2.3) for each control strategy (HL, RA and RL) from all the fifteen participants.

Figure 4.4 (a) and (b) present the resultant human force and tracking error over the Whole trajectory, i.e. from L1, L2, L3, L4 to L1. Figure 4.4 (c) and (d) present the resultant human force and tracking error over the first half of the trajectory with obstacles, i.e. from L1, L2 to L3. Figure 4.4 (e) and (f) present the resultant human force and tracking error over the second half of the trajectory without obstacles, i.e. from L3, L4 to L1.

Figure 4.5 shows the average human force and tracking error of the fifteen participants at each timestep during the experiment (Metrics (3) and (4) in Section 6.2.3).

4.6.1 Average Magnitude of Human Force

1. **Over the Whole Trajectory**: Figure 4.4(a) shows that the RA control strategy can reduce the magnitude of human force by 44.9% and 40.6% compared to strategies of HL and RL, respectively.
2. **With obstacles (over the trajectory from L1, L2 to L3):** Figure 4.4(c) shows that RA reduces the magnitude of human force by 44.1% and 51.6% compared to HL and RL, respectively.
3. **Without obstacles (over the trajectory from L3, L4 to L1):** Figure 4.4(e) shows that RA reduces the magnitude of human force by 45.6% and 10.3% compared to HL and RL, respectively.

4.6.2 Average Tracking Error

1. **Over the Whole Trajectory**: Figure 4.4(b) shows that RA reduces the tracking error by 18.4% compared to HL and increases 10.3% compared to RL.
2. **With obstacles (over the trajectory from L1, L2 to L3):** Figure 4.4(d) shows that RA reduces the tracking error by 9.60% compared to HL and increases 4.44% compared to RL.
3. **Without obstacles (over the trajectory from L3, L4 to L1):** Figure 4.4(f) shows that RA reduces the tracking error by 39.7% compared to HL and increases by 38.5% compared to RL.

4.6.2.1 Discussion of Human Force and Tracking Error

It was realized that the force applied by humans under the RL control strategy is significantly higher than that of RA and HL over the segment of the trajectory with obstacles (i.e. from L1, L2 to L3). This may be related to the human trying to take control to avoid the obstacles, which can be observed in Figure 4.5.

The tracking error from RL is the smallest because the robot is in control during the experiment and the position of the moving target is known to the robot. However, the obstacles positions are not known to the robot. Therefore, the robot will be likely to move toward the obstacles. This has been shown in table 4.1 that RL has the highest failure rate, among the three control strategies, of 46.5%. It can be seen in Figure 4.5(a), that the variation of the human force in RL is largest compared to HL and RA because the

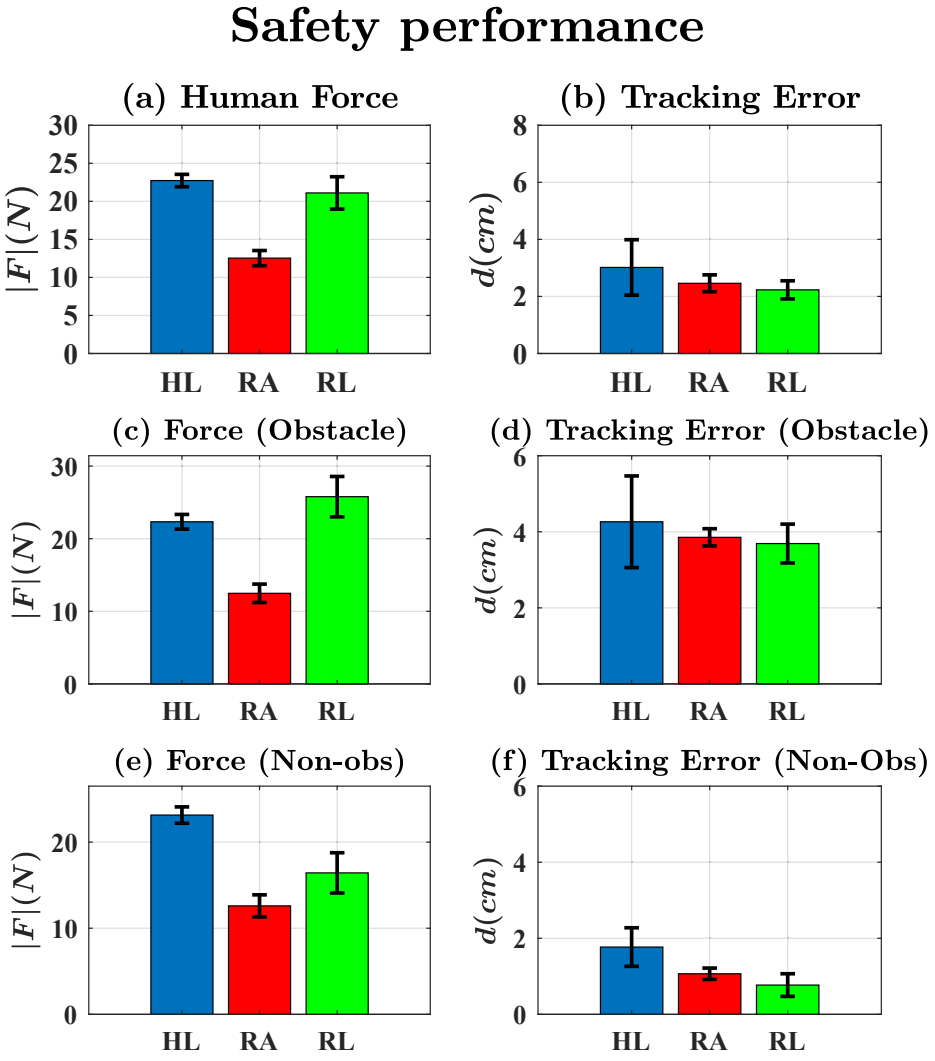


FIGURE 4.4: Evaluations of the three control strategies from the fifteen participants: (a) Human Force (Overall) (b) Tracking Error (Overall) (c) Human Force (Obstacle) (d) Tracking Error (Obstacle) (e) Human Force (Non-obstacles) (f) Tracking Error (Non-obstacle). Obstacles: First half of the trajectory with obstacles, i.e. from L1, L2 to L3. Non-obstacle: Second half of the trajectory without obstacles, i.e. from L3, L4 to L1.

human co-worker needs to apply extraordinary force to take control when closing to the obstacles. Therefore, the gradient of interaction force may be an indicator of whether the human co-worker wants to be in control or signals a low interaction experience.

An interesting observation from the study is the highest standard deviation (std) of magnitude of force observed during RL implementations, as detailed in Figure 4.4(a)(c)(e). This variation highlights the inconsistent reliability of autonomous control across different individuals, suggesting a correlation between an individual's inclination towards trusting autonomous systems and their resultant magnitude of force to show intention of taking

control. Specifically, participants with a higher degree of trust in autonomous control had a lower of magnitude of force applied. It can be hypothesized that such variability is largely attributable to differing levels of trust and dependence on the autonomous system among participants.

Furthermore, when examining tracking errors, the standard deviation associated with HL is noted to be the highest as shown in Figure 4.4(b)(d)(f). This outcome is anticipated, given the absence of autonomous guidance in influencing the robot's movements, thereby introducing variability directly tied to the participants' diverse skill levels. This is exemplified by the tracking error standard deviations for the RI-cohort (proficient group) at 2.22cm and for the Non-RI-cohort (non-proficient group) at 4.14cm, further underscoring the impact of individual participant skill on performance outcomes.

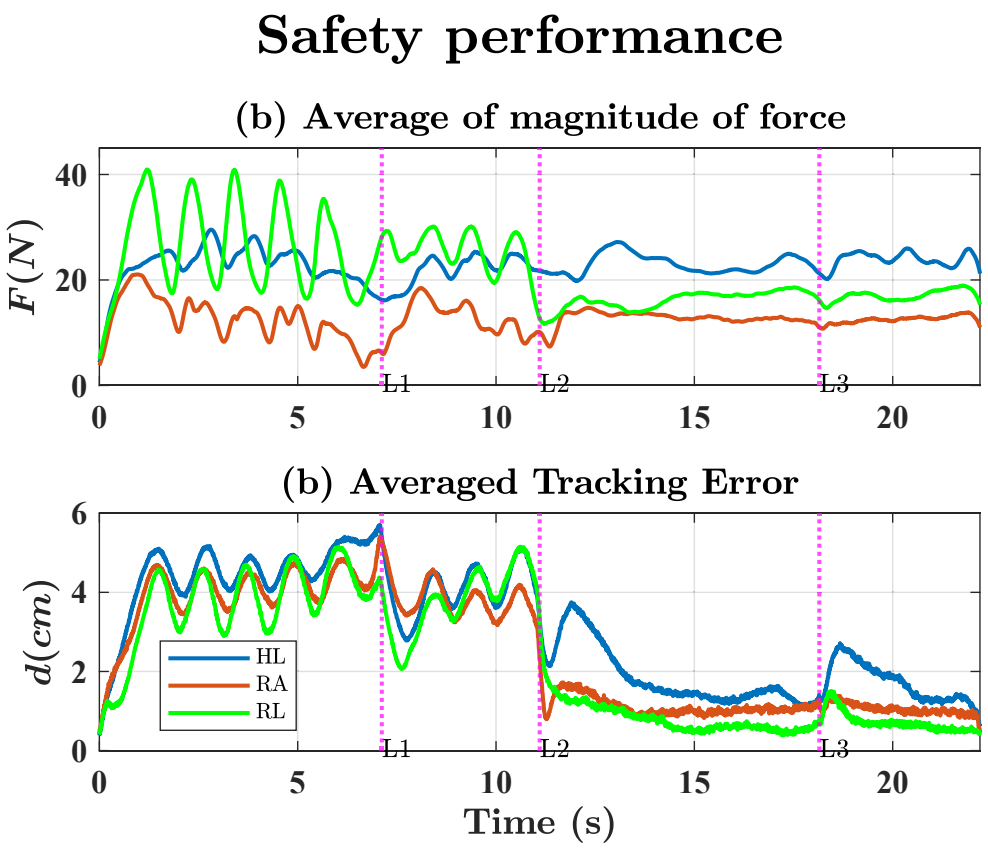


FIGURE 4.5: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants over the experiment at each timestep. The three pink dotted vertical lines are the time step referring to L2, L3 and L4 in Figure 4.3.

TABLE 4.1: Failure rate for the three control strategies

Control Strategies	HL	RA	RL
Failure Rate (%) (FR)	5.3	12.7	46.5

4.6.3 Failure Rate

Table 4.1 shows the failure rate of the three control strategies. It was realised that the failure rate of RA is similar to HL due to the participant may not pay full attention to the task. And the participants do not become familiar with robots at the beginning and may collide with the obstacles accidentally. When familiarity increases, the failure rate decreases. Moreover, RL has the highest failure rate among the three control strategies.

4.6.4 NASA TLX

Figures 4.6 and 4.7 show the results of the questionnaire based on both the rating scale (RS) and pairwise comparison (PC) methodologies, measuring participants' perceived workload through the NASA TLX framework. NASA TLX is used for participants to provide their feedback. It has six dimensions, including mental demand, physical demand, temporal demand, performance, effort, and frustration. Low scores represent low demand on the user, and are therefore preferred.

The alignment between the RS and PC results can be attributed to the marked disparities among RA, RL, and HL, a finding consistent with the conclusions drawn in Chapter 5. In particular, Figures 6.19 and 6.20 underscore RA's superior performance over RL and HL, the discrepancy being particularly pronounced for RL ($RL >> HL > RA$). It can be seen that RA is the lowest (best performance) and RL is the highest (worst performance) in the six dimensions among all control strategies as shown in Figure 4.6. It can also be seen that RA is the highest (best performance) and RL is the lowest (worst performance) in the six dimensions among all control strategies as shown in 4.7. The significant difference between RA and RL is caused by the robot being misled when closing on obstacles, which are reported by the participants. Therefore, the interaction experience is significantly reduced due to the considerable conflict between humans and robots when close to obstacles. An interesting result is that the perceived physical demand is the largest in the RL as shown in Figure 4.6(b). However, the human forces of HL are the highest (Figure 4.4(a)). Consequently, perceived physical demand is not only affected by the objective physical demand applied to the robot, but may also be mental and frustration levels. When frustration and mental demand are high, perceived physical demand may also increase correspondingly. This can be an interesting research direction that can be further investigated in the future.

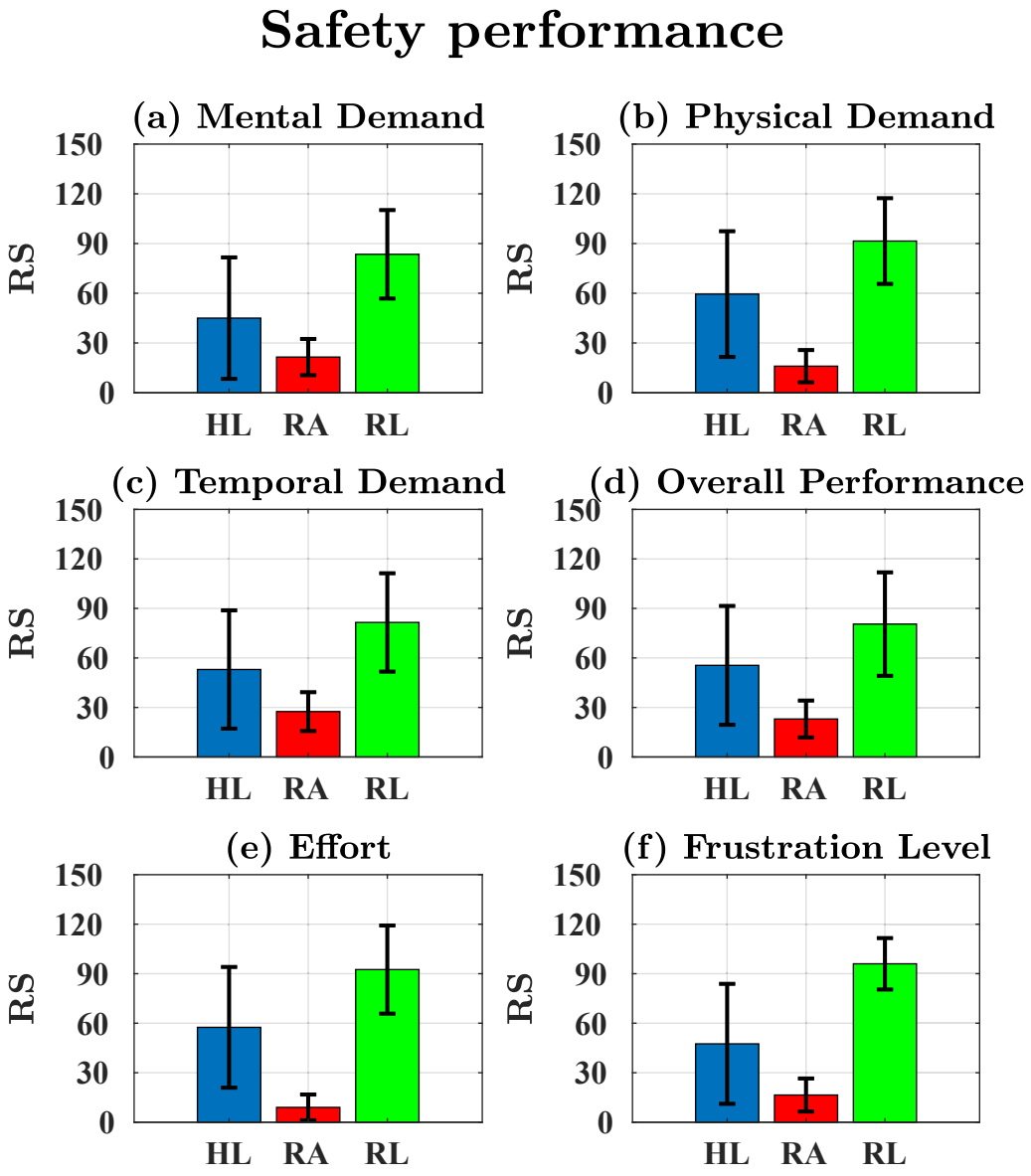


FIGURE 4.6: Normalized subjective results from NASA-TLX with rating scale method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants. 0 - good performance, 100 - bad performance.

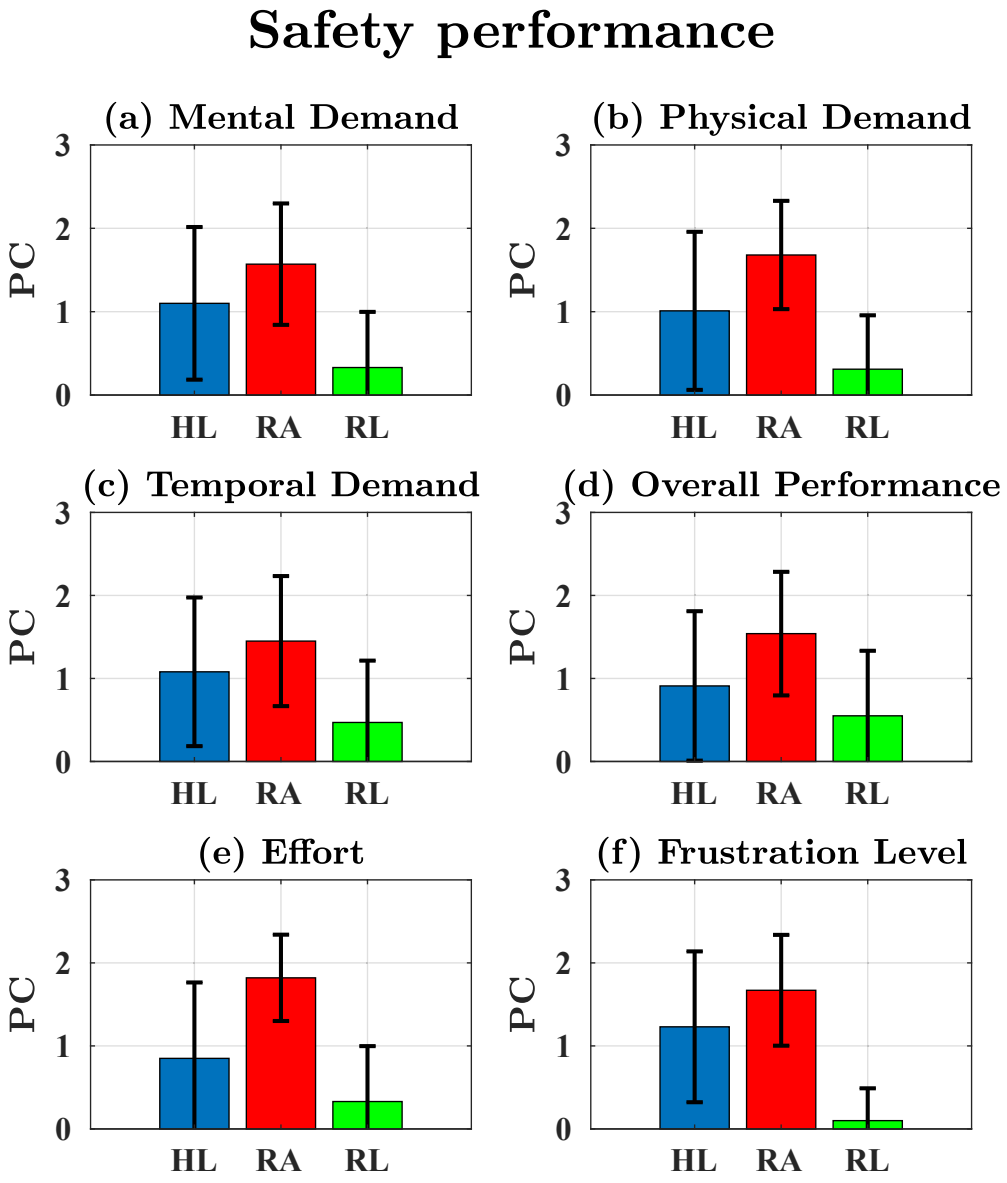


FIGURE 4.7: Subjective results from NASA-TLX with pairwise comparison method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants.

4.7 CONCLUSION

A TSC-based role arbitration method was proposed and experimentally validated with the human subjects' experiment. The results show that the RA control strategy can reduce the magnitude of human force by 44.9% and 40.6% compared to the HL and RL strategies,

respectively. Furthermore, RA reduces the tracking error by 18.4% compared to HL and increases by 10.3% compared to RL. However, RL has the highest failure rate, among the three control strategies, at 46.5%. In terms of subjective evaluation, RA reduces 67.3% and 79.8% compared to HL and RL, respectively. In conclusion, the proposed role arbitration method has been shown to achieve superior combined human-robot performance, reduce the workload of human colleagues and improve subjective preference.

In exploring the computational intricacies of robot trust, self-confidence, and their respective roles in decision arbitration, one underlying objective emerges as critical: approximating human-human interaction. The closer robots can emulate and align with human behaviours and expectations, the more fluid and natural their collaborative interactions will become. Thus, the foundation of this collaboration rests on understanding the human side of the interaction equation. When striving for more human-like interactions, gauging the subtleties of human perception and experience is essential. Although robots' arbitration methods, blending trust and self-confidence, showcase technical sophistication, the ultimate measure of their success lies in the quality of interaction they deliver. With this perspective, the discussion transitions to the next chapter, emphasising the qualitative aspects of these interactions.

Chapter 5

Measuring Subjective Impression during Physical Human-Robot Collaboration Using a Pairwise Comparison Method

5.1 Introduction

In earlier discussions, the foundational principles of robot trust in humans, robot self-confidence, and the role arbitration method were highlighted, all centred around a core objective: making human-robot interactions feel akin to human-human interactions. Although these technical and theoretical advancements are important, human experience remains paramount.

This realisation prompted the following exploration of alternative methods. In this chapter, the focus is on pairwise comparison (PC) method, a novel survey approach designed to address the aforementioned challenges to address sub-research question 4. The pairwise method offers a more intuitive approach. Instead of asking participants to rate items on an abstract scale, this method asks them to compare two items directly. Such a comparative framework can often feel more engaging and straightforward. Participants can access and articulate their feelings more easily when comparing two tangible options, which makes it less likely for them to respond carelessly. Thus, this direct comparison can capture more genuine and accurate subjective impressions. The pairwise method has not really been used before as a method of evaluating user preferences in pHRI. Through this method, we

aim to capture human subjective impressions with enhanced precision, offering a clearer window into the nuanced world of human-robot interactions.

To compare the proposed pairwise method with the more traditional rating scale method, a series of pHRC experiments is conducted that compare the results of these two methods to the ground truth. This thesis chose the NASA TLX as the Likert scale assessment tool because of its simplicity, popularity, and validated accuracy. The PC questionnaire is designed by modifying the questions in the NASA TLX. This thesis employed the one-way analysis of variance (ANOVA). Moreover, Tukey's HSD multiple comparison test was also applied to allow us to determine specifically between which groups the differences existed. Through the results, it is concluded that the PC method is more robust, efficient, and reliable than the Likert scale method in determining user's impressions in pHRC studies, especially when the difference between two comparing group's subjective impression is subtle regardless of the variations in experimental procedure designs. The experimental results have been promising, highlighting a reduction in the experimental time 44%. Even within this shortened timeframe, the pairwise method proved to be more adept at capturing genuine subjective impressions, remedying many of the issues we previously faced with traditional rating scales. Furthermore, participants reported that they found the questionnaire process easier and more pleasant, marking a significant step forward towards bridge theory and lived experience in human-robot collaboration studies.

The remainder of this chapter is structured as follows: Section 5.2 presents the experimental protocol. In Section 5.3 the experimental results are presented and extensive analysis is provided. Finally, the discussion and conclusion are presented in Section 5.4 and Section 5.5 respectively.

5.2 Experiment Protocol

5.2.1 The Clock Game

The Clock Game is shown in Fig. 5.1. The position of the cobot end effector is projected onto the TV through coordinate frame transformation and is represented by the white dot. The red dot is the moving target. It starts from the top left corner of the yellow rectangular trajectory and moves in a clockwise direction at constant speed. The task of the experiment is to control the white dot to follow the red moving target as closely as possible. To ensure the participation and focus of the participants during the experiment, a competitive element is added to the task. A circular boundary is designed to move synchronously with the target dot along the trajectory. The experimental subjects are asked to keep the white dot within the radius of the circular boundary as its size shrinks

over time, making the task more challenging and requiring higher concentration. Based on how long the participant maintains the white dot within the boundary, the *Score* is calculated in real time at a rate of $45Hz$ and displayed to the participant. At each time step, while the white dot is kept within the boundary, a point is assigned to the *Score*. On the contrary, the *Score* will be penalised by 10 points if it is outside the limit at each time step.

FIGURE 5.1: The experimental task requires participants to follow the red target dot with the white dot while remaining within the red circular boundary. The task becomes more intense as the radius of the boundary reduces with time. A *Score* is calculated based on the time that the participant could stay within the boundary. It increases if the participant is able to remain within the boundary. Otherwise, it reduces.

5.2.2 Collaborative Robot (Cobot) Setup

The cobot used in this experiment is the ANBOT which is a robot system of physical collaboration designed for industrial abrasive blasting [115]. This platform was used in Chapters 3 and 4 during the development and verification of the trust model and TSC-based arbitration models.

The ANBOT consists of the 6-DoF Universal Robot UR10 manipulator with a custom-made handle at its end effector that incorporates a 6-DoF force/torque sensor. The ANBOT is controlled by a mass-damping admittance controller [116] as shown in (5.1). $M_d \in \mathbb{R}^{6\times6}$ and $D_d \in \mathbb{R}^{6\times6}$ are the virtual inertia and damping matrix respectively; $\dot{x} \in \mathbb{R}^{6\times1}$ is the desired velocity of the end effector in the Cartesian space and $\ddot{x} \in \mathbb{R}^{6\times1}$ is the desired acceleration; $F \in \mathbb{R}^{6\times1}$ is the collaborative wrench applied to the cobot end-effector by the human operator, it is measured directly by the force/torque sensor.

$$M_d\ddot{x} + D_d\dot{x} = F + \tau \tag{5.1}$$

The term $\tau \in \mathbb{R}^{6\times 1}$ is the artificial noise deliberately injected into the admittance controller to influence the perception of the human operator in a predicable manner. Details are provided in the next subsection.

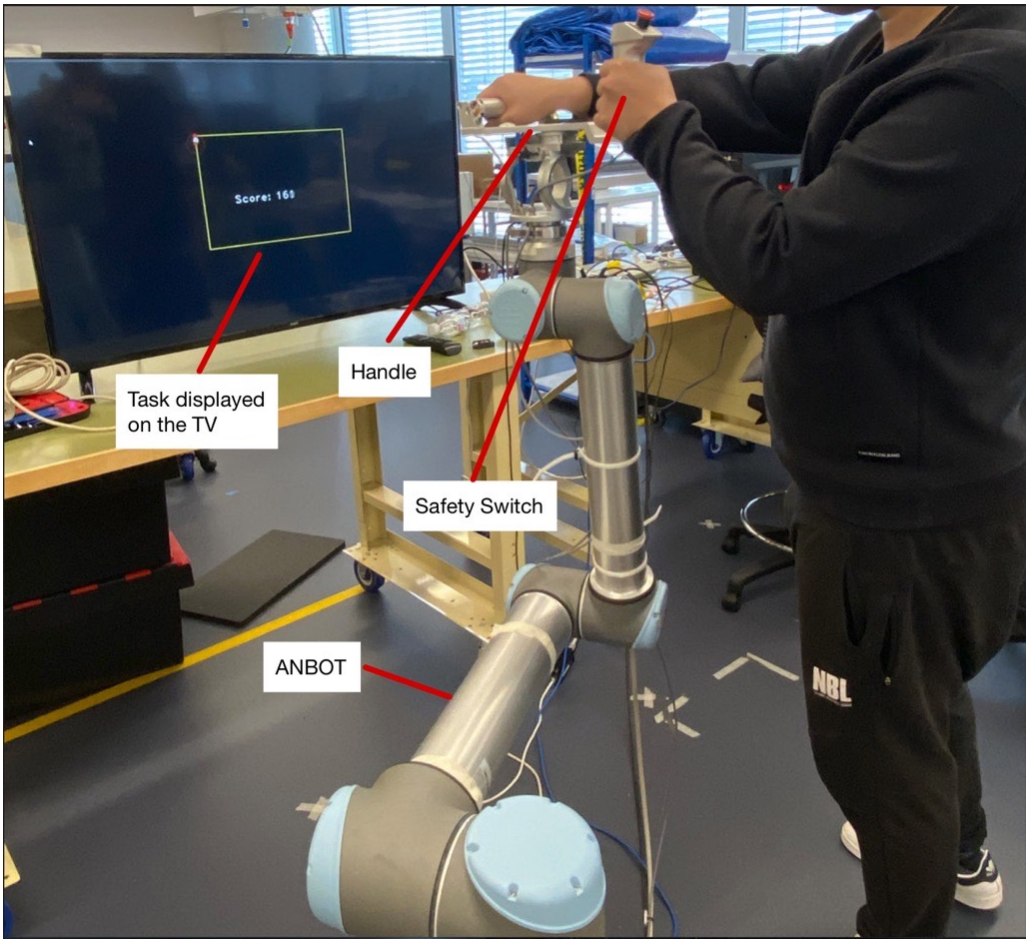


FIGURE 5.2: A participant conducting the experiment using the custom-made handle attached at the end-effector of ANBOT. The safety switch is used to allow the participant to activate ANBOT and control the start of the experiment.

Considering the Clock Game task introduced in the section above, the ANBOT is restricted to planar motions. For F and τ, only the elements that contribute to the planar motion are used by the admittance control law (5.1).

5.2.3 Artificial Noise Design

In Equation 5.1, the artificial noise τ is used to establish the ground truth for the experiment based on the hypothesis that noise added to F will deteriorate the user experience: the larger the noise is, the worse the experience should be.

τ is randomly generated every 0.5s and injected into a random quadrant of the plane of allowable movement. The magnitude of τ is randomized following the Gaussian distribution between the lower limit L and the upper limit U. Four noise levels are determined through a pilot study based on the principle that the difference between two successive noise levels is so subtle that a pHRC expert cannot distinguish them based on trials.

- τ_1*:* L=0N, U=0N. *No disturbance.*
- τ_2*:* L=1N, U=2N. *Little disturbance.*
- τ_3*:* L=3N, U=4N. *Medium disturbance.*
- τ_4*:* L=5N, U=6N. *Large disturbance.*

Throughout the entire experience, the noise levels used are kept unknown to participants.

5.2.4 Questionnaires Design

The NASA TLX questionnaire is used. With considerations of the experiment length and the relevancy of the original NASA TLX questions to our experiment design, the following two questions and the standard 21-point scale [88] are used in the experiment:

- *RS Q1: How physically demanding was the task?*
- *RS Q2: How frustrated were you? For example, were you insecure, discouraged, irritated, stressed or annoyed?*

For the PC questionnaire, we created the following questions by referring to the NASA TLX questionnaire but adapting them into a pairwise comparison format.

- *PC Q1: Which mode required less physical demand?*
- *PC Q2: Which mode was less frustrating to use?*

Participants are asked to compare the two noise levels given to them on a 5-point scale and points are assigned to the noise levels accordingly:

- **A≫B:** The performance of A is much better than that of B. 2 points for A, 0 points for B.

- **A>B:** The performance of A is a little better than that of B. 1 point for A, 0 point for B.
- **A=B:** Performance of A and B is about the same. 0 points for both A and B.
- **A<B:** B performance is slightly better than A. 0 point for A, 1 point for B.
- **A≪B:** B performance is much better than A. 0 points for A, 2 points for B.

The first noise level given to the participant is marked as A and the subsequent one is marked as B.

5.2.5 Experimental Procedure

Two experimental procedures are designed which correspond to conducting all the experimental conditions (Experimental procedures 1 and 2). Both the RS and PC questionnaires are carried out in the following experiments.

5.2.5.1 Experimental Procedure 1

For the first experimental procedure, participants need to complete the task for all combinations of noise level comparisons, that is: τ_1&τ_2, τ_1&τ_3, τ_1&τ_4, τ_2&τ_3, τ_2&τ_4 and τ_3&τ_4 with single-order matchups (e.g, A versus B and no B versus A). Using a balanced Latin square [117], six distinctive sequences of conducting the comparisons are generated to reduce the order effect.

For the NASA TLX questionnaire, participants take trials of the noise levels in the sequence generated by the balanced Latin square above and give scores after every trial.

From this experimental procedure, twelve NASA TLX results and six pairwise comparison results are collected from each participant. The order in which the questionnaires are completed is also alternated. For example, if the previous participant completes the NASA TLX questionnaire first, the subsequent participant will receive the PC questionnaire first.

5.2.5.2 Experimental Procedure 2

In the second experimental procedure, the experimental sequence generated in Experimental Procedure 1 is used likewise. Except, participants are only required to conduct the experiment 4 times for either questionnaire instead of 12. As a result, only 2 PC questionnaire results and 4 NASA TLX results are collected. The order in which to complete the questionnaires is also altered.

5.2.6 Hypothesis

It is reasonable to anticipate that participants' subjective sensations for physical demand and frustration will be ranked from the best to the worst in line with the magnitude of the noise: τ_1 should result in the least physical demand and frustration, followed by τ_2, τ_3 and τ_4 in sequence. Based on this presumption, the following hypotheses are made for different experimental procedures:

- *Hypothesis 1*: For Experimental Procedure 1, both questionnaire methods should return the correct result that matches the anticipation. NASA TLX results should show that τ_1 has the lowest scores for physical effort demand and frustration, with the scores increasing sequentially for τ_2, τ_3, and τ_4. For the PC results, τ_1 should receive the most points, followed by τ_2, τ_3 and τ_4 in descending order.
- *Hypothesis 2*: For Experimental Procedure 2, it is hypothesised that only the PC results will match the anticipation. NASA TLX will fail to rate the noise levels correctly because of inter-rater variations and insufficient data from each participant.

5.2.7 Participant Recruitment:

The experiment was carried out with the approval of the Human Research Ethics Committee of the University of Technology Sydney (UTS) ETH18-3029. Participants are recruited through personal connections and incidental engagements. Participants are given more details about the experimental task through an information sheet and verbal explanation. A demonstration of the experiment is also provided. After this, participants are given the consent form to read and sign. Before starting the experiment, the questionnaire questions are shown to the participants and they will practice with robot under τ_1 (no artificial noise) until the score reaches 500 as shown in Figure 5.1 to guarantee the same level of proficiency in starting the experiment to reduce the learning effect between individuals.

5.3 Results

The experiment was carried out on the UTS Open Day (OD) and Orientation Week (OW) with an age group between 16 and 40 years. During these events, a cohort of 36 inexperienced volunteers (16 women and 20 men) who had no prior experience with pHRC was recruited. Nine of them followed Experimental Procedure 1 and the rest performed Experimental Procedure 2. Furthermore, 18 volunteers with pHRC experience (2 women and 16 men) were recruited at the UTS Robotics Institute (RI) in the weeks following and

were split into half for the experimental procedures. For the 18 participants who completed Experimental Procedure 1, their first 8 experimental runs and the corresponding questionnaire results are extracted and combined with the results of Experimental Procedure 2. For either experimental procedure, the balanced Latin square was completed in full 3 times.

For NASA TLX results, the mean and standard deviation of each noise level are calculated from the participant-rated scores directly, and the results are presented in Fig.5.3a and Fig.5.4a for Experimental Procedure 1 and Experimental Procedure 2, respectively. From the results of the PC questionnaire, the noise levels points are obtained from the comparison results as per Section 5.2.4. The mean and standard deviation of the points are calculated, and the results are shown in Fig. 5.3b and Fig. 5.4b for two experimental procedures, respectively.

A ANOVA is performed to test whether there was a statistically significant difference in the means between different noise level scores. The p-value tests whether the true difference of means for two noise levels is equal to zero, in other words, whether two noise levels resulted in similar levels of physical demands and frustration in our experiment. Typically, a p-value < 0.05 indicates a significant difference between the means of two groups [118]. The results are shown in Table 5.1.

5.3.1 Hypothesis 1: Both RS and PC will return agreeing and statistically significant results in Experimental Procedure 1

Fig. 5.3 shows the results of Experimental Procedure 1 and is used to test Hypothesis 1. RS mean scores are presented in Fig.5.3a. For both Physical Demand and Frustration, RS returned a result that agreed with our anticipation. The average scores are rated the lowest for τ_1 and start to climb for τ_2, τ_3 and τ_4. This indicates that participants found τ_1 the least physically demanding and resulted in the least frustration, followed by τ_2, τ_3 and τ_4 in sequence. From the second and fourth columns of Table 5.1, the p-values for τ_1 vs τ_3, τ_1 vs τ_4, τ_2 vs τ_3 and τ_2 vs τ_4 can be viewed as less than 0.05, indicating that RS showed statistically significant differences between these noise levels in the categories of physical demand and frustration. However, the p-values for τ_1 vs τ_2 and τ_3 vs τ_4 are all *greater* than 0.05 (cells highlighted in yellow) which are statistically insignificant, implying that within these pairs, participants found these noise levels resulted in similar levels of physical demand and frustration.

As can be seen in Fig. 5.3b, PC has returned perfectly agreeing results for Physical Demand and Frustration. The average points obtained by τ_1 for the two tested criteria are

substantially higher than τ_2 by 25.6% and 23.9%, respectively. The average points obtained by τ_3 are much lower compared to τ_1 and τ_2, and they are approaching 0 for τ_4. These reviews that participants found τ_1 resulted in the least physical demand and frustration, followed by τ_2, τ_3, and τ_4 in sequence. To test the statistical significance of the PC results, p-values in the third and fifth columns of Table 5.1 should be studied. As can be seen, all the p-values in these columns are all lower than the standard significance level. For the noise level pairs where RS failed to differentiate (τ_1 vs τ_2 and τ_3 vs τ_4), PC has successfully shown the statistically significant differences between them (cells highlighted in green).

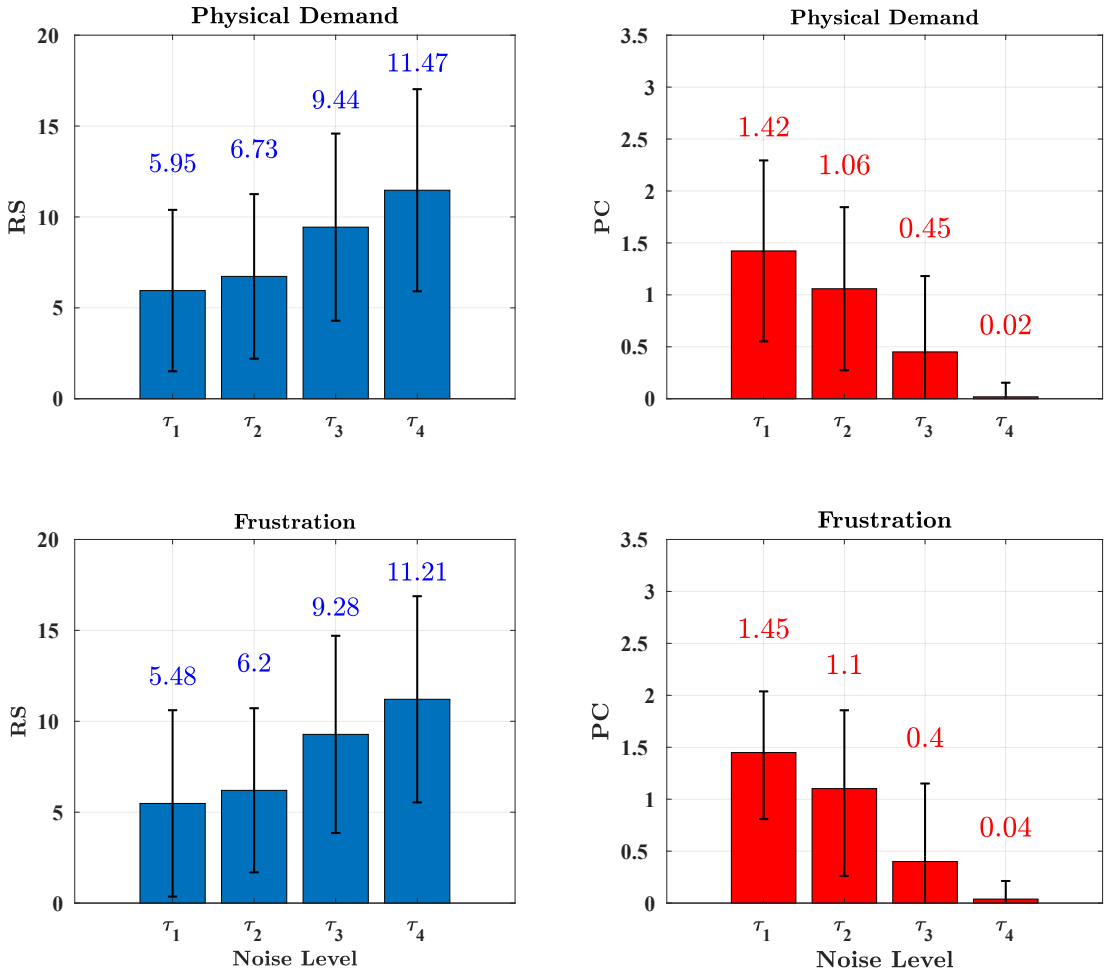
These results provide evidence to reject the null hypothesis 1: The RS results agree with our anticipation shown in Figure 5.3a. However, it is not statistically significant. However, PC was capable of returning statistically significant results that matched our anticipation.

5.3.2 Hypothesis 2: PC can return statistically significant and agreeing results under Experimental Procedure 2 and RS will fail

The Experimental Procedure 2 results are shown in Fig. 5.4 and they are used to test Hypothesis 2. From Fig. 5.4a, it can be observed that RS has performed poorly for both the physical demand and the frustration categories. For the first, the error occurred between τ_1 and τ_2, the RS average scores show that the participants found τ_1 more physically demanding than τ_2. For frustration, noise levels were rated from the best performance to the worst performance on the order of τ_2, τ_1, τ_4 and τ_3. These results violate the results of Experimental Procedure 1 and our anticipation. Furthermore, the sixth and eighth columns of Table 5.1 show statistical significance. Similarly to the results of the Experimental Procedure 1, the mean values are tested to be statistically insignificant for τ_1 vs τ_2 and τ_3 vs τ_4.

However, as shown in Fig.5.4b, PC persisted in relating participants' perceptions with noise levels correctly. The smallest noise level achieved the highest average point; as larger and larger noises were used, lower and lower points were obtained by them. Furthermore, the PC results are validated to be statistically significant according to columns 7 and 9 of Table 5.1. All the p-values are less than the standard significance level, including the noise levels with subtle differences (τ_1 vs τ_2 and τ_3 vs τ_4 as highlighted in green).

Experimental Procedure 2 shows that RS fails to return the correct and statistically significant results when the number of repetitions for the experiment is limited for each participant. At the same time, the PC accuracy lasts. Therefore, Hypothesis 2 should be accepted.

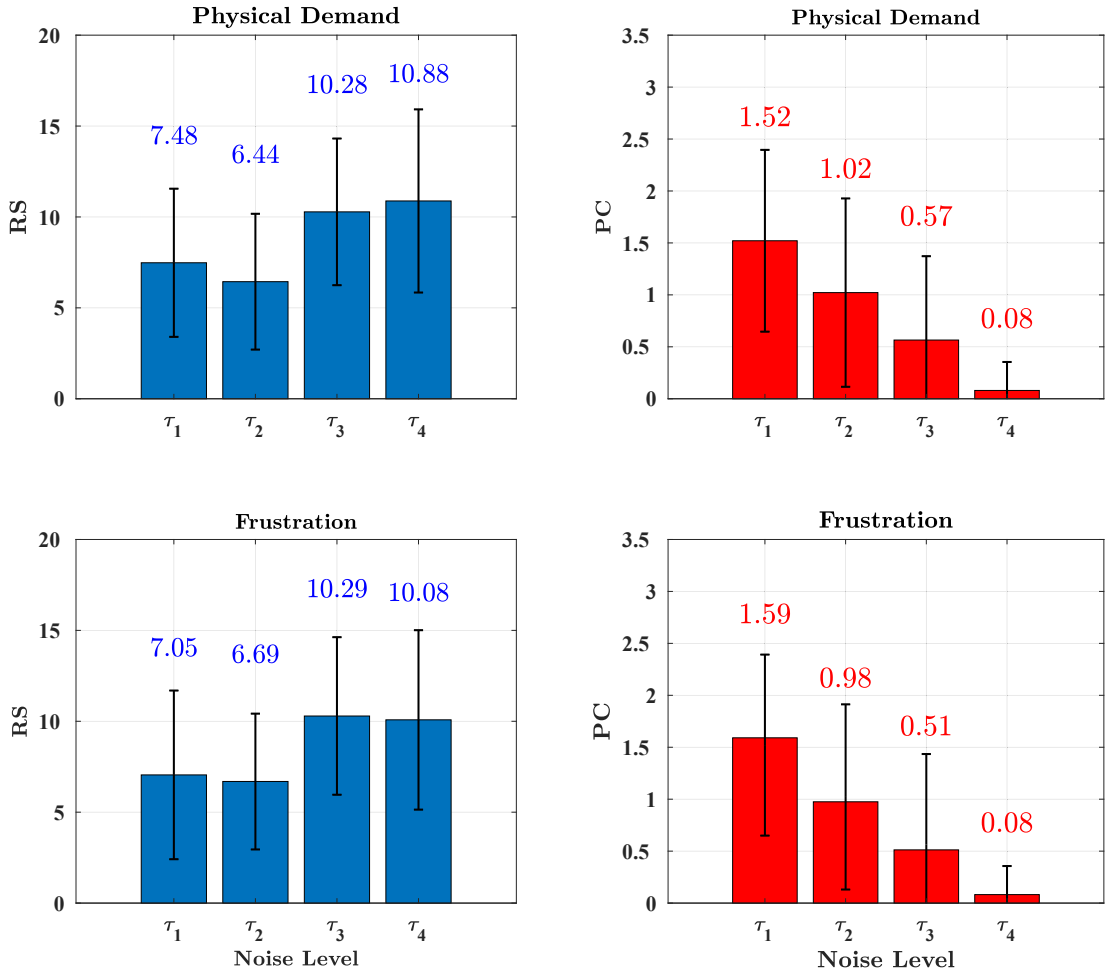


(a) Rating Scale (b) Pairwise Comparison

FIGURE 5.3: Experimental Procedure 1 results for (a) Rating Scale Questionnaire (b) Pairwise Comparison Questionnaire

5.3.3 Experimental Duration

In Experimental Procedure 1, participants spent an average of *25 minutes* to complete all the NASA TLX questionnaires. It should be noted that this length included 3 repetitions of the noise levels. The average completion time of the PC questionnaire is *14 minutes*.



(a) Rating Scale (b) Pairwise Comparison

FIGURE 5.4: Experimental Procedure 2 results for (a) Rating Scale Questionnaire (b) Pairwise Comparison Questionnaire

TABLE 5.1: Comparison of rating scale and pairwise comparison method by employing the p-value from the one-way ANOVA followed by Tukey's HSD multiple comparison tests for experimental procedures 1 and 2.

Noise Level	Experimental Procedure 1 Physical Demand		Experimental Procedure 1 Frustration		Experimental Procedure 2 Physical Demand		Experimental Procedure 2 Frustration	
	RS	PC	RS	PC	RS	PC	RS	PC
τ_1 vs τ_2	0.822166	0.037120	0.872085	0.027475	0.575070	0.007382	0.972555	0.001132
τ_1 vs τ_3	0.000503	0.000000	0.000304	0.000000	0.002536	0.000000	0.000531	0.000000
τ_1 vs τ_4	0.000000	0.000000	0.000000	0.000000	0.000063	0.000000	0.000955	0.000000
τ_2 vs τ_3	0.013759	0.007265	0.006597	0.000000	0.000014	0.019474	0.000112	0.030415
τ_2 vs τ_4	0.000001	0.000000	0.000001	0.000000	0.000000	0.000000	0.000202	0.000000
τ_3 vs τ_4	0.113456	0.007265	0.179922	0.016261	0.862868	0.008956	0.993072	0.036399

5.4 Discussion

In this thesis, the use of the pairwise comparison approach was used in pHRC studies and a preliminary study was conducted to compare PC results with the more popular RS method. Through testing our hypotheses, the following key advantages of PC over RS can be identified:

1. When the tested characteristics are repeated multiple times, both RS and PC mean values provide accurate measurements of participants' subjective impressions. However, only PC results are statistically significant in determining differences for characteristics with subtle differences. RS results are not statistically significant.

2. When the duration of the experiment is reduced, the PC mean values accurately rate human subjective feelings and are statistically significant. RS mean values are noisy and fail to reflect human's true subjective perceptions.

A potential cause of the poor RS results may be the absence of a mutual understanding of the rating system between participants. The individual participant must establish their own rating standards by retrospectively comparing the current result with their previous responses when conducting any rating scale questionnaire [119]. When the number of repetitions of the experiment is limited, as in Experimental Procedure 2, the experimental subjects do not establish the rating standard. Therefore, they provided noisy and inaccurate results using RS. Furthermore, cultural influences [39] [40], the length of the RS scale [38] and careless responses may have also contributed to the poor RS results.

A limitation associated with PC is the complexity of the experiment design. To compare the characteristics n, PC requires a total of $(n-1)^2/2 + (n-1)/2$ comparisons for a single order match-up comparison while RS requires only a minimum of n experiments. However, as validated in Experimental Procedure 2, the accuracy of PC does not depend on individual participants completing all comparisons. Robustness is particularly useful for pHRC studies because the experimental task is often repetitive and experimental subjects will become fatigued both mentally and physically after many repetitions. PC reduces the workload of the experimental subjects more effectively than RS. However, the increased number of experiments required by PC leads to higher recruitment costs because more experimental subjects are needed.

TABLE 5.2: Comparison between Rating Scale and Pairwise Comparison methods

	Rating Scale	Pairwise Comparison
Effectiveness with repetition (Experiment 1)	No	Yes
Effectiveness without repetition (Experiment 2)	No	Yes
Averaged Experimental length	25 minutes	**14 minutes**
Robustness	Low	**High**
The level of satisfaction in answering questions	Low	**High**
The level of satisfaction in conducting the experiments	Low	**High**
Recruitment cost	**Low**	High

5.5 Conclusion

In this thesis, a novel comparison is conducted between two subjective impression evaluation methods. The classic NASA-TLX questionnaire was used as a template to create the Rating Scale and Pairwise Comparison questionnaires, respectively. Through two experimental procedures of different lengths, the effectiveness of RS and PC was statistically compared. The experimental results show statistically significant results for both RS and PC when the difference between the comparing groups is large. However, when the true difference is small, PC is more robust and accurate than RS regardless of the experimental procedure. Furthermore, compared to RS, PC also reduced the duration of the experiment by 44%, resulting in a more enjoyable experimental experience and lower cognitive workloads.

A summarised comparison between the Rating Scale and the Pairwise Comparison is shown in Table 5.2. Even though the cost of recruitment is higher for PC, it still offers a more time-efficient, accurate, and robust alternative method to pHRC researchers who are interested in capturing nuances in human perceptions during pHRC studies.

The subsequent chapter delves into a comprehensive experimental analysis using the TSC role arbitration method to validate these dimensions in human-robot collaboration. The PC method validated in this chapter will be used in the next chapter to determine the user impressions of the TSC role arbitration method.

Chapter 6

Experimental Evaluation of Role Arbitration Method for Physical Human-Robot Interaction

6.1 Introduction

While the previous chapters laid the foundation for understanding human-robot collaboration through the TSC-based role arbitration method, the previous empirical endeavours have provided just a glimpse into its full potential. This chapter builds upon the work presented in earlier sections, amalgamating the methods of the TSC-based role arbitration and evaluating the system with a series of experiments involving human participants.

Figure 1.10 shows the framework of the TSC-based role arbitration method based on the theoretical contribution from Chapters 3 and 4. Previous explorations, as outlined in Chapter 4, centred exclusively on the safety aspect of the computational model of robot trust in humans. The current chapter broadens this scope, integrating components of the computational robot trust model into the TSC role arbitration framework and emphasising three critical performance factors: smoothness, physical, and cognitive. Singularity, although crucial, is left aside for this evaluation, given the extensive research conducted by Carmichael [2, 111] on combined performance and subjective impressions during pHRC.

Safety, singularity, and smoothness have been identified as factors that affect human performance in pHRC [2, 99, 111, 120]. Yet, comprehensive evaluations regarding physical and cognitive performance in the domain remain limited. This chapter addresses this research gap.

Subsequent sections investigate the validation of Physical and Cognitive Performance, followed by the design and execution of experiments assessing the efficacy of the TSC-based role arbitration method. Later sections introduce evaluation metrics that include the magnitude of force, tracking error, and subjective impressions, comparing them with human- or robot-centric control.

Furthermore, this chapter also embarks on a comparative study between the pairwise and rating scale methods for all the performance factors, including smoothness, physical, and cognitive, leaning on the NASA-TLX questionnaires as previously discussed. This juxtaposition helps capture the subtle intricacies of subjective impressions on robots during pHRC.

The findings in this chapter highlight the effectiveness of the TSC-based role arbitration method. The proposed method not only accentuates the combined performance of humans and robots but also significantly decreases the physical and cognitive workloads on the human counterpart. In tandem, the subjective impressions of humans towards robots in a collaborative setting are enhanced, underscoring the tangible impact of this theoretical concept.

6.2 Experimental Evaluation

6.2.1 Experiment Testbed

For the work presented in this chapter, the same experiment testbed will be used as previously presented in Section 3.4.1 in Chapter 3.

6.2.2 Design of Experiment

Physical human-robot collaboration can take advantage of human and robot strength, human decision making, and robot accuracy and strength because the autonomous system is not capable of performing all tasks, especially in unstructured environments. In a structured environment, the autonomous system can be more capable than human co-worker due to its inherent nature of accuracy and power. However, in an unstructured environment, human intervention is often necessary to address the flaws of an autonomous system. One of the flaws in the autonomous system is the perception system. Due to sensor noise, surrounding light, occlusions, or dusty industrial environment, the autonomous system may often not be able to perform the desired task. Therefore, an experiment similar to that described in Section 3.4.2 but based on this scenario is designed.

A red-filled-circle moving target (RMT) tracking experiment is designed to verify the proposed TSC-based role arbitration method as shown in Figure 6.1. In this experiment, a human co-worker is asked to move the robot end-effector to track the RMT. To simulate a realistic and challenging industrial application, noise is added to the pose of the tracking target (noise is generated according to Equation 6.1). This noise represents the inability of the robot to accurately determine the true location of the RMT, replicating a scenario such as a system with poor perception due to sensor error. In this work we refer to this scenario as an *unstructured* environment. Therefore, the position of RMT will be perceived differently between humans and robots (p_{RMT} and p^N_{RMT}) from L1-L2 and L2-L3 as shown in Figure 6.1. From L3-L4 and L4-L1, the environment is *structured* without this disturbance, therefore, there is no sensor noise. Consequently, the pose of RMT is the same from the perspective of the human and robot (p_{RMT} and p_{RMT}).

In order to investigate how each performance component (Smoothness, Physical, and Cognitive) affects the performance role arbitration method, the task requirement (i.e., high tracking accuracy), state of RMT (i.e., position, and speed of RMT) are manipulated in the following experiments.

$$p^N_{RMT} = p_{RMT} + \mathcal{N}(0, \sigma^2) \tag{6.1}$$

The parameter σ is the standard deviation of the noise. p^N_{RMT} is the pose of RMT with noise, with a new noise value computed every 1 second. p_{RMT} is the pose without noise.

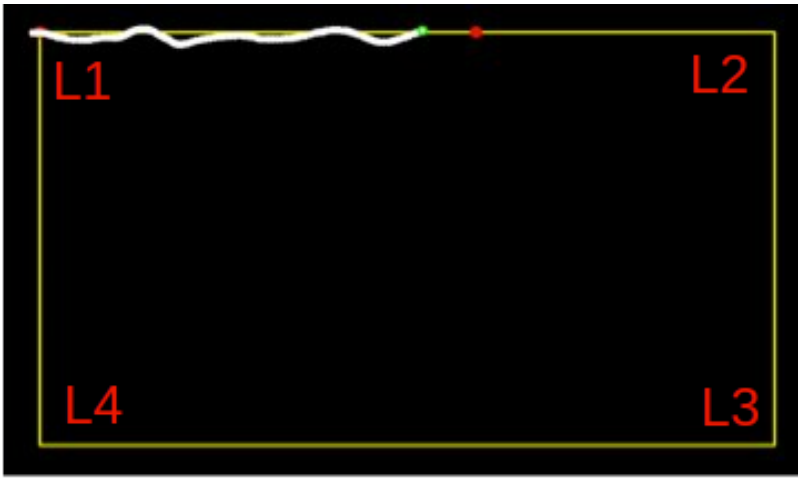


FIGURE 6.1: Experiment Scenario: The red-filled-circle moving target (RMT) starts from the top left corner (L1) and moves around the rectangle path (yellow line) in a clockwise direction with constant speed. The green-filled circle is the current end-effector position. When the moving target returns to the starting position (L1 → L2 → L3 → L4 → L1), it is regarded as completing one loop. The white line is the actual trajectory of the robot end-effector. L1, L2, L3 and L4 are the timestep when the RMT reaches four corners of the rectangle path.

6.2.3 Three Control Strategies and Evaluations

For comparison, three arbitration control strategies are implemented in the experiments:

1. *HL - Human Leading*: Human in control $\alpha = 0$ (Equation 4.1)
2. *RL - Robot Leading*: Robot in control $\alpha = 1$ (Equation 4.1)
3. *RA - Role Arbitration*: Robot Trust and Self-Confidence Based Role Arbitration Method $\alpha(TSC)$ (Equation 4.11)

To evaluate the effectiveness of the proposed arbitration method (RA) compared to fixed arbitration (HL and RL), the following metrics are used:

1. *Average Magnitude of Human Force Over the Whole Trajectory*: $\frac{1}{Nc}\sum_0^{Nc}(\frac{1}{t_N - t_1}\sum_{t_1}^{t_N}||F_H(t)||)$. Nc is the total number of experiments for the corresponding control strategies. t_1 is the initial timestep of the experiment. t_N is the final timestep. $||F_H(t)||$ is the magnitude of human force at each timestep.
2. *Average Tracking Error Over the Whole Trajectory*: $\frac{1}{Nc}\sum_0^{Nc}(\frac{1}{t_N - t_1}\sum_{t_1}^{t_N}||e(t)||)$. $||e(t)||$ is the tracking error which is the distance between the position of the end-effector and the moving target at each timestep.
3. *Average Magnitude of Human Force at each timestep*: $\frac{1}{Nc}\sum_0^{Nc}||F_H(t)||$.
4. *Average Tracking Error at each timestep*: $\frac{1}{Nc}\sum_0^{Nc}||e(t)||$.
5. *Failure Rate*: $FR = \frac{\text{number of Failed experiments}}{\text{Total number of experiments}}$. Failure experiments refer to the situation when the tracking error is more than a threshold value (The threshold value is different based on different experiment design which will be discussed in each results and discussion Sections).
6. *NASA TLX*: Post-experiment questionnaires for measuring the perceived workload of the participants, which includes mental demand, physical demand, temporal demand, performance, effort, and frustration.

6.2.4 Participants and Testing Procedure

Fifteen subjects participated in the experiment. Ten participants are from the UTS Robotics Institute and five participants are from the general public. The ages of the participants range between 20 and 40. After completing the consent form, each participant learns to operate the robot (shown in Figure 3.7a) until they are familiar with the

task. This is to reduce the learning effect. After these steps, each participant will perform five tests for each control strategy in a random order to reduce the effect of order. Participants are asked to answer the questionnaires once they have completed the experiments. This experiment follows the procedure approved by the Ethics Committee of the UTS with approval number ETH21-6346.

6.3 Experiment 1 - Smoothness Performance

6.3.1 Pre-experiment Parameter Identification

sm^- and sm^+ are task-dependent parameters for Equation 3.11, which are used to specify a level of smoothness in the operator's movements that is consistent with good performance. For a task that requires high accuracy, such as a surgical task, both sm^- and sm^+ are set to small values. For speed-demanding tasks, such as speed-demanding pick and place tasks where smoothness may not be critical, sm^- and sm^+ can be set large.

A quantitative way to obtain these parameters is to determine them from data collected from experts performing the task. Expert data ($sm[1:T]$) is collected from an expert conducting the experiment.

$$\begin{cases} sm^- = max(sm[1:T]) \\ sm^+ = \gamma \times max(sm[1:T]) \end{cases} \tag{6.2}$$

γ is determined by how strict the smoothness requirement is. If the task has a strict smoothness requirement, then γ should be set to smaller ($\gamma \to 1$).

6.3.2 Experiment Design

In order to investigate how smoothness performance affects the performance of the role arbitration method, the tracking accuracy requirement is expected to be high for participants.

Based on Equations 6.2, $\gamma = 2$, $sm^- = 1.5$, and $sm^+ = 3$. This was determined from data collected from 5 experts. Experts are from the UTS Robotics Institute, which has a long period of experience in using the ANBOT. Each taking 5 trials of the defined task. The table below shows the experimental parameters for experiment 1.

TABLE 6.1: Smoothness Experimental Parameters

Experiment Parameter	Value
High Tracking Accuracy	Yes
Speed of RMT (cm/s)	7.5
Random movement of RMT (cm)	a = 0, b = 0
Number of Loops	1

6.3.3 Results and Discussion - Experiment 1 Smoothness Performance

Figure 6.2 shows the mean and standard deviation of the human co-worker's average tracking error and magnitude of force (Metrics (1) and (2)) for each control strategy (HL, RA and RL) from all fifteen participants. Figures 6.2 (a) and (b) present the resultant human force and tracking error on the whole trajectory, that is, from L1, L2, L3, L4 to L1. Figures 6.2 (c) and (d) present the resulting human force and tracking error in the first half of the trajectory, which has sensor noise on the desired red moving target in Figure 6.1, that is, from L1, L2 to L3. Figures 6.2 (e) and (f) present the resulting human force and the tracking error over the second half of the trajectory without sensor noise, i.e. from L3, L4 to L1.

Figure 6.3 shows the averaged magnitude of the force and the tracking error at each timestep as shown in Figure 6.3(a) and (b) (Metric (3),(4)).

Table 6.2 shows the failure rate of HL, RA and RL based on the criteria that when the tracking error exceeds the 10 cm threshold (Metric (5)).

Figures 6.4 and 6.5 show subjective measurement of the impression of the participant based on the rating scale (RS) and the pairwise comparison method (PC) in Chapter 5 (Metric (6)).

6.3.3.1 Average Magnitude of Human Force

1. **Over the Whole Trajectory:** Figure 6.2(a) shows that the RA control strategy can reduce the magnitude of human force by -36.90% and -8.16% compared to strategies of HL and RL, respectively.

2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.2(c) show that RA reduces the magnitude of human force by -28.50% and -20.81% compared to HL and RL, respectively.

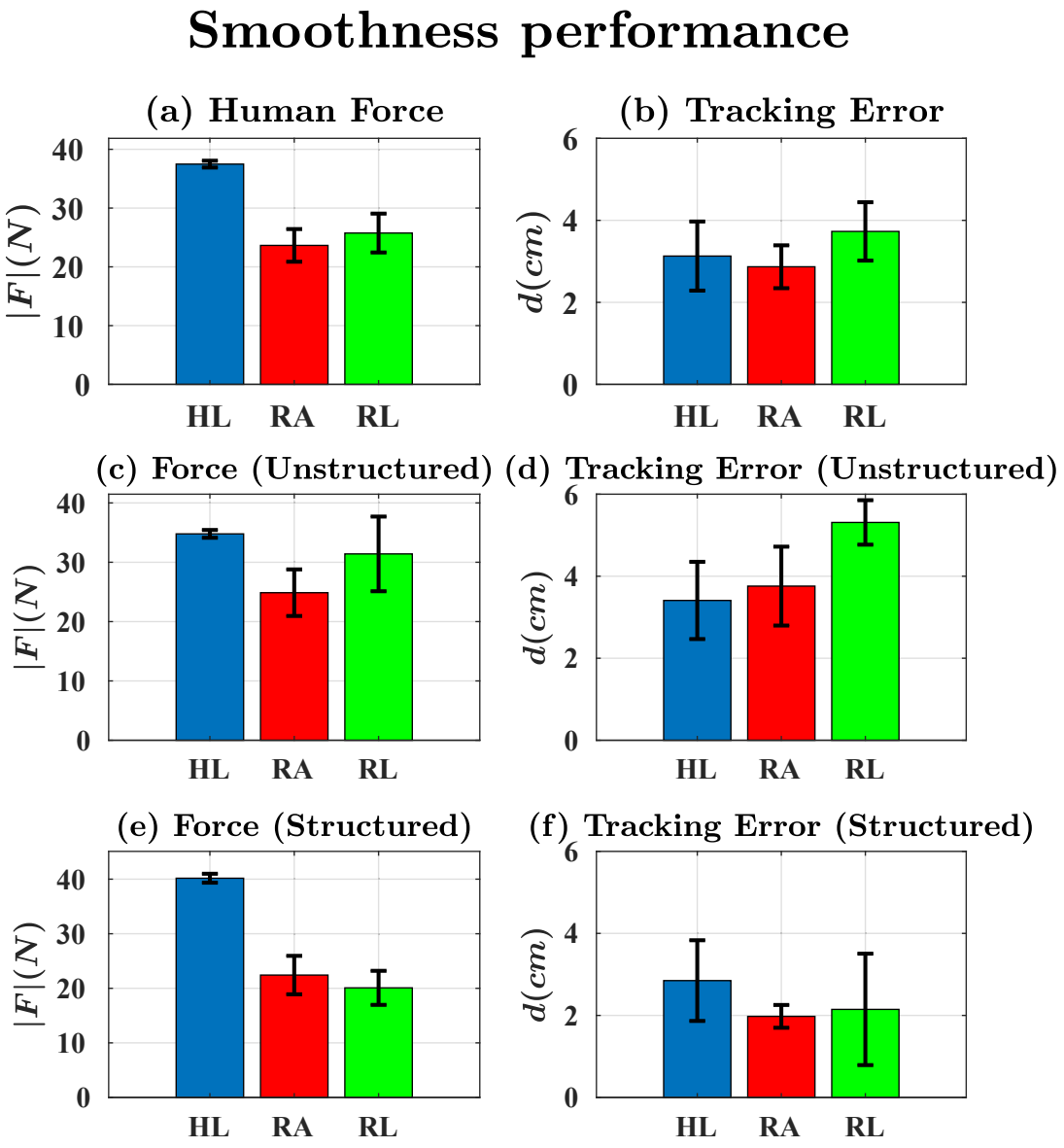


FIGURE 6.2: Evaluations of the three control strategies from the fifteen participants for smoothness performance in Experiment 1 in Section 6.3 based on the evaluation criteria 1, 2 in Section 6.2.3: (a) Human Force (Overall) (b) Tracking Error (Overall) (c) Human Force (Unstructured) (d) Tracking error (Unstructured) (e) Human Force (Unstructured) (f) Tracking error (Unstructured). Unstructured: The first half of the trajectory with sensor noise on the desired red moving target in Figure 6.1, i.e. from L1, L2 to L3. Structured: Second half of the trajectory without sensor noise, i.e. from L3, L4 to L1. The error bar is the standard deviation.

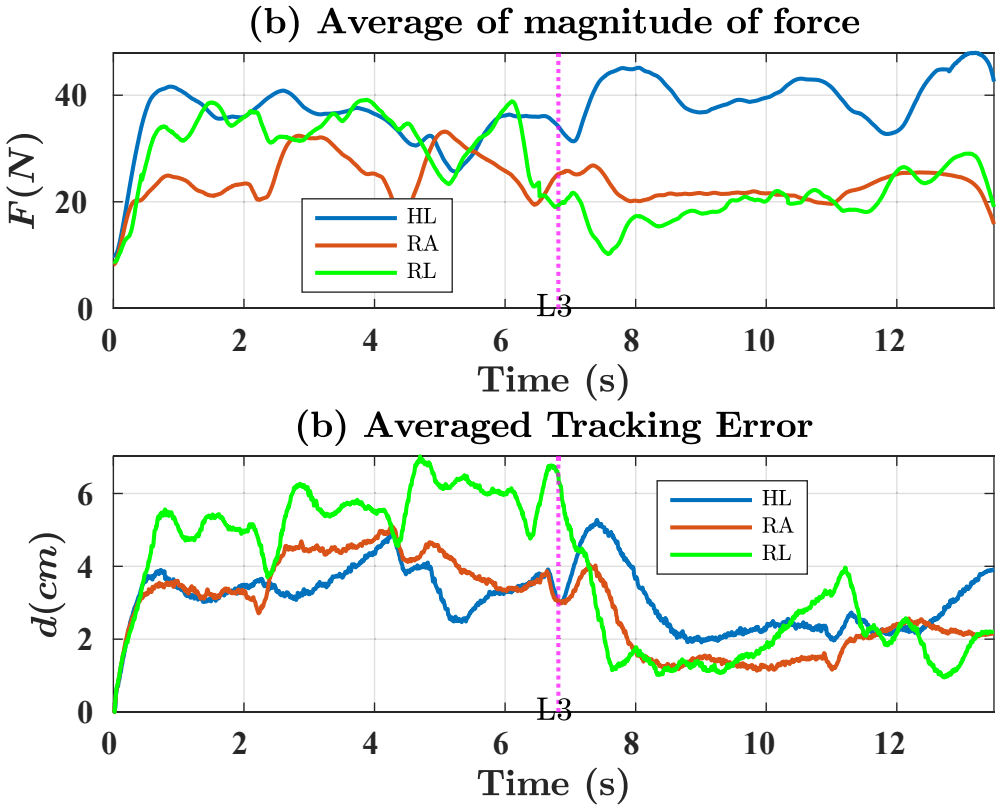


FIGURE 6.3: Evaluations of the three control strategies from the fifteen participants for smoothness performance in Experiment 1 in Section 6.3 based on the evaluation criteria 3, 4 in Section 6.2.3: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants over the experiment at each timestep. The pink dotted vertical lines are the time step referring to L3 in Figure 6.1. The left side of L3 corresponds to the unstructured environment with sensor noise and the right side of L3 corresponds to a structured environment.

3. **Structured (over the trajectory from L3, L4 to L1):** Figure 6.2(e) show that RA reduces the magnitude of human force by -44.2% compared to HL, but increased by +11.7% compared to RL.

6.3.3.2 Average Tracking Error

1. **Over the Whole Trajectory:** Figure 6.2(b) shows that RA reduces the tracking error by -8.33% compared to HL, and increases +23.10% compared to RL.

2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.2(d) shows that RA reduces the tracking error by -10.25% compared to HL, and increases +29.23% compared to RL.

3. **Structured (over the trajectory from L3, L4 to L1):** Figure 6.2(f) shows that RA reduces the tracking error by -30.06% compared to HL, and increases by +7.92% compared to RL.

6.3.3.3 Discussion of Human Force and Tracking Error

The findings align with the anticipation that autonomous control excels in a structured environment ($RL < HL$ in Figures 6.2(e)(f), represents lower magnitude of force undertaken and less tracking error introduced for RL), whereas human co-workers are more adept in unstructured settings ($HL < RL$ in Figures 6.2(c)(d)). In Figure 6.2(c), it is evident that RL outperforms RA. However, when corresponding to a structured environment in Figure 6.2(e), the trend reverses with $RL < RA$. This principle remains consistent when observing the tracking error in Figures 6.2 (d) and (f). Here, the order is $HL < RA < RL$, which can be attributed to the autonomous system being misled by the sensor noise at the position of RMT. Similarly, in Figure 6.3, a notable observation is that the force exerted by humans under the RL control strategy considerably exceeds that of RA and HL, especially during the transition through the unstructured environment, specifically on the left-hand side of L3. This trend is likely a consequence of the unpredictable and errant movements initiated by the robot control. Consequently, human co-workers apply additional force to regain control. In addition, the worse performance caused by misguided autonomous control results in a significant decrease in subjective impression, which can be observed in Figure 6.4 and Figure 6.5. These results demonstrate that the proposed role arbitration method artfully integrates the strengths of human and autonomous controls. The result is a noticeable improvement in the combined performance of human-robot collaborations (lower tracking error) and a decrease in the physical workload carried out by the human co-worker (lower human force) in general in (a) and (b) in Figure 6.2.

Another intriguing discovery is the elevated standard deviation (std) of the tracking error during RL in structured settings in Figure 6.2(f), which underscores the variable reliability of autonomous control between individuals. Those inclined to trust autonomous control might lean on it more, resulting in superior tracking accuracy. Conversely, individuals sceptical of the robot's capabilities might oppose its actions, leading to subpar performance. A hypothesis can be made that this variance can be largely attributed to the different levels of trust and reliance on the autonomous system among participants according to the participant's self-reporting and based on the observations. When analysing tracking errors, the std of HL emerges as the highest. This is expected since there is no autonomous intervention shaping the robot's movements, leading to discrepancies arising from participants' varying skill levels due to RI-cohort (profieint group) is 1.94cm, and

Non-RI-cohort (Non-profieint group) is 3.98cm. Therefore, an interesting research direction can investigate the dynamic of human trust in robot and robot self-confidence, as well as the level of reliance on autonomous control when the human co-worker interacts with the proposed role arbitration method. A larger and more comprehensive trust dynamic between humans and robot could be formulated to combine the proposed robot trust in human co-worker and robot self-confidence model with human trust in robot and human trust in itself model. Based on the comprehensive trust dynamic, a new role arbitration method could be investigated and derived.

6.3.3.4 Failure Rate

The failure rate represents the bottom line for the successful completion of the task. If the participant exceeds more than a certain tracking error, which is considered as the required task performance for the current experiment design, then the corresponding experiment is considered a failure. The delineated threshold tracking error for Experiment 1 is anchored at 10 cm, as indicated in Table 6.2.

It should be noted that RL with a failure rate of 43.75% significantly overshadows both HL and RA, both of which register at 0%. This substantial difference is primarily due to the inherent challenges posed by the unstructured environment, as shown in Figure 6.3. In such conditions, the movements directed by the robot control can be erratic and unpredictable due to the addition of sensor noise, and with the robot leading the interaction, there is reduced ability for the human co-worker to promptly intervene and adjust the end effector's pose. In contrast, for both RA and HL modes, the robot's actions are considerably more foreseeable due to the robot movement is totally or partially under control by the human co-workers. This predictability ensures that participants find it more manageable to steer the robot and maintain the tracking error within the set threshold.

Table 6.2: Smoothness experiment failure rate of HL, RA and RL based on the evaluation criteria 5 in Section 6.2.3

Threshold: 10cm	HL	RA	RL
Failure Rate %	0.00	0.00	43.75

6.3.3.5 Subjective - NASA-TLX

Figures 6.4 and 6.5 showcase the questionnaire results, capturing the perceived workload of the 15 participants. These evaluations are rooted in the NASA-TLX framework, with data gathered via the rating scale (RS) and pairwise comparison (PC) methodologies.

NASA TLX offers participants a structured platform to articulate their feedback across six axes: mental demand (MD), physical demand (PD), temporal demand (TD), performance (PER), effort (EFF), and frustration (FRU) as shown in (a) - (f), respectively.

An examination of the results underscores the similarity between the RS and PC evaluation methods, especially in the way that they capture the six-dimensional measurements from the NASA-TLX. This congruence aligns well with the revelations from Chapter 5, which, when the difference between the control strategies is significant, the results obtained through the RS and PC methods are similar. In terms of interaction experience, RA stands out as the most favourable, followed by HL, and then RL. Indeed, the erratic and unpredictable behaviour of the robot in the RL setting puts it at a disadvantage ($RL >> HL > RA$). This unpredictability manifests itself even when objective metrics do not significantly distinguish between RA, RL, and HL. A noteworthy observation is in the realm of physical workload: objectively, HL records the highest strain as shown in Figure 6.2(a)(c)(e), but subjectively, this does not seem to dominate participants' impressions. A hypothesis is made that this disconnect is likely because participants are primarily attuned to the unpredictability inherent in the robot's movements, particularly evident in the conflicting forces that emerge between humans and robots in unstructured settings.

Feedback from participants also suggests a noticeable trend: there is a discernible reluctance to utilise the robot under the RL mode. Consequently, in the design of robotic systems with human-in-the-loop considerations, it is imperative to strike a balance. Although task performance remains undeniably crucial, the role arbitration paradigm underscores the need to prioritise subjective impressions. If an autonomous control system is adept at tracking but is poor at identifying the correct target, it can foster more significant misalignment, leading to heightened conflict in interaction experiences. Conversely, if the system's tracking capabilities fall short, its performance leaves much to be desired, particularly in structured environments. These insights reinforce the argument that, while task performance is vital, designers must give great consideration to the subjective impressions of users when crafting robotic systems with human elements.

6.3.3.6 Conclusion - Smoothness Performance

The one-way analysis of variance (ANOVA) performed shows the statistically significant difference ($p < 0.05$) for all the evaluation metrics in Section 6.2.3 between HL, RA, and RL. In conclusion, the proposed role arbitration method has been shown to achieve superior combined human-robot performance, reduce the workload of human coworkers, and improve subjective preference when incorporating smoothness performance measures.

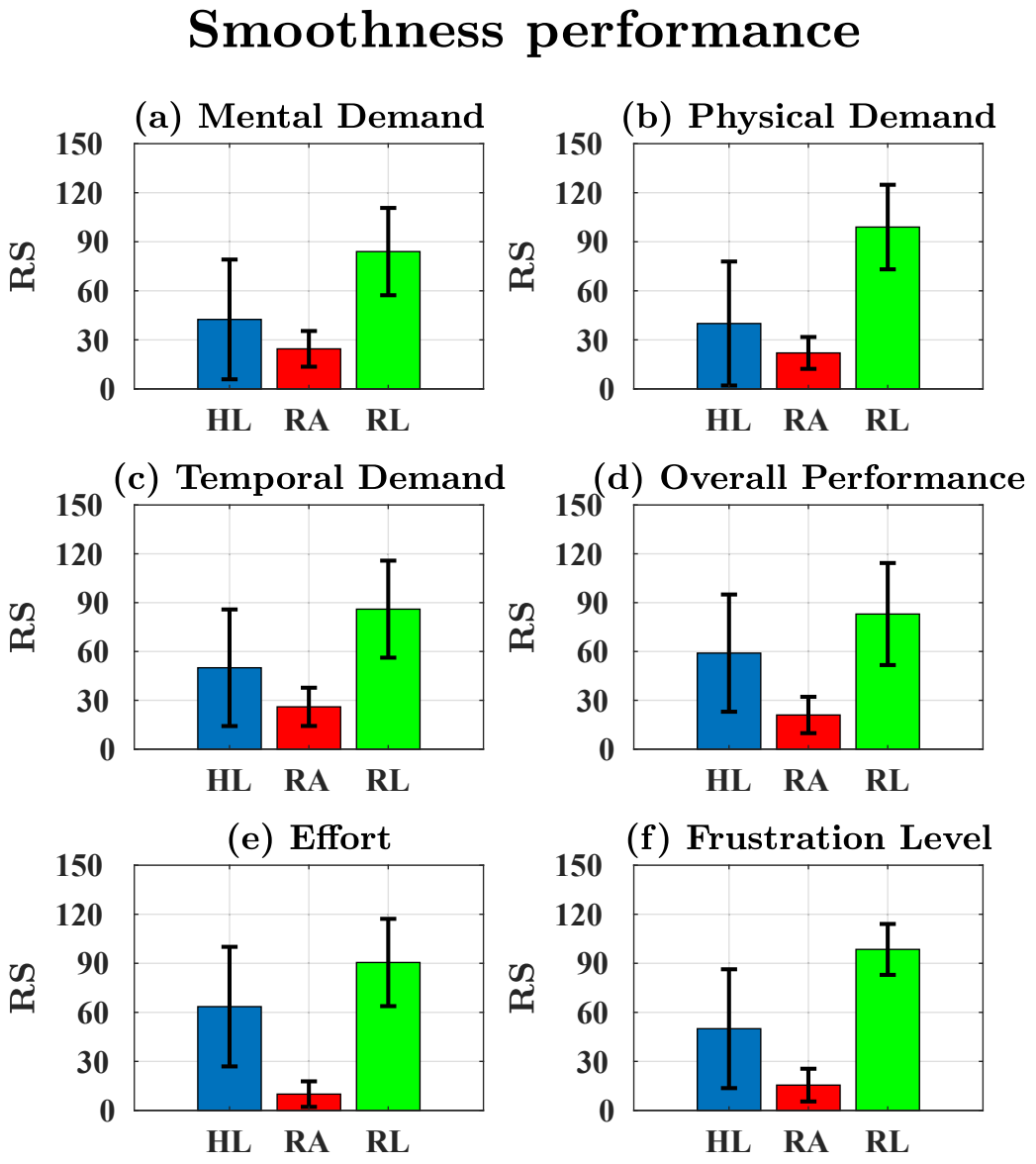


FIGURE 6.4: Normalized subjective results from NASA-TLX with rating scale method (RS) based on Evaluation Criteria 6 (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants for the smoothness experiment. 0 - good performance, 100 - bad performance.

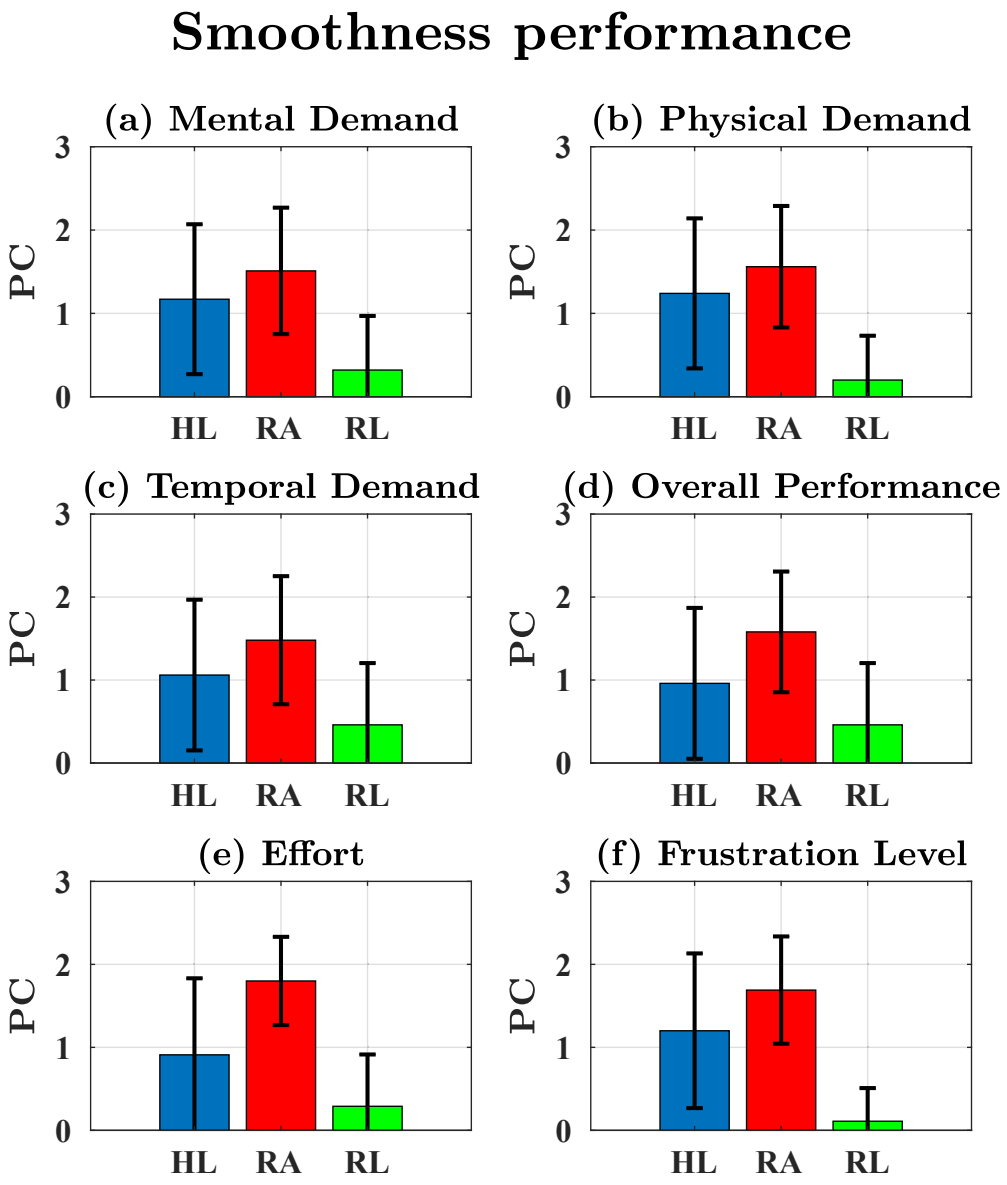


FIGURE 6.5: Subjective results from NASA-TLX with pairwise comparison method (PC) based on Evaluation Criteria 6 (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants for the smoothness experiment. 0 - bad performance, 2 - good performance.

6.4 Experiment 2 - Physical Performance

6.4.1 Pre-experiment Parameter Identification

Modelling physical performance accurately requires a nuanced understanding of how it varies between tasks and individuals. Factors such as the specific muscle groups engaged during a task or whether the task is of high or low workload play a pivotal role. As such, this section presents a pre-experiment parameter identification process tailored to capture the intricacies of physical performance. In order to identify the physical performance

FIGURE 6.6: Hand Dynamometer used for measuring maximum isometric force.

parameters fatigue coefficient, recovery coefficient, and maximum voluntary contraction (Cf, Cr, MVC) in Equation 3.12 and 3.13, a pre-experiment test is designed. Participants are asked first to perform a maximum effort task, followed by a rest experiment. The maximum isometric force (MVC) is measured through a hand dynamometer [121] as shown in Figure 6.6 because MVC is a standard method to measure the strength of the upper limb [122]. At first, we collect the data as MVC ($F_{max,iso}[0] = MVC$). Next, participants are asked to move the robot as hard as possible to ensure maximum effort in their muscle groups, so that we keep the external force in the muscle groups required for the task maximum ($F_{ext} = F_{max,iso}$) for 4 minutes. During this task, we measure $F_{max,iso}$ three times every 30 seconds through a hand dynamometer and calculate the average of the three values. After the exercise session, a 4-minute rest is performed, during which we measure $F_{max,iso}$ at 30-second intervals in the same way as previously described. During the rest period, the external load is zero ($F_{ext} = 0$). Therefore, based on a sequence of measured $F^M_{max,iso}$, an optimisation method is formulated using fmincon in MATLAB as shown in Equation 6.3:

$$\min_{Cr,Cf,MVC} \quad \sum_{n=1}^{N} (F^M_{max,iso}[n] - F_{max,iso}[n])^2 \tag{6.3}$$

The measured $F^M_{max,iso}$ and the curve fitting $F_{max,iso}$ is shown in Figure 6.7.

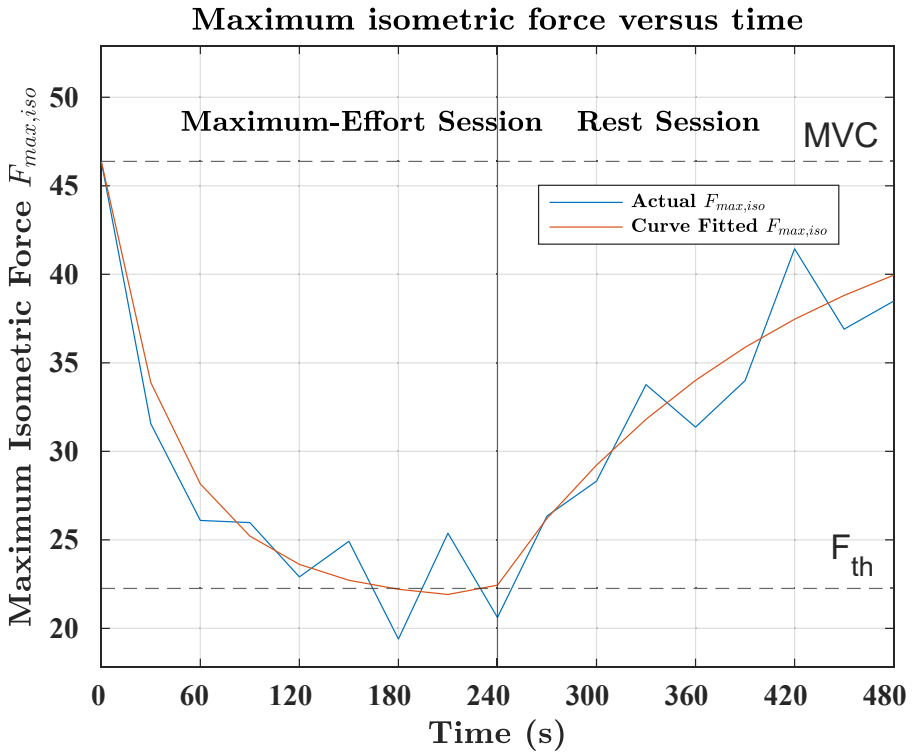


FIGURE 6.7: One of the participant pre-experiment parameter identification: Maximum isometric force measured $F^M_{max,iso}$ through the hand dynamometer and fitted curve $F_{max,iso}$ (Equation 6.3). The left side of the four-minute cut-off line (vertical black line) is the maximum-effort session and the right side is the rest session. The dashed horizontal line MVC is the maximum isometric force. The identified physical performance based on the curve fitted result is $MVC = 45, Cf = 0.012, Cr = 0.006$.

6.4.2 Experiment Design

In order to impose physical fatigue on the participants, the speed of RMT is designed to be large ($V_{RMT} = 15cm/s$) as shown in Table 6.3. The trajectory of the RMT adheres to the anticipated path, thereby resulting in a minimized cognitive workload.

TABLE 6.3: Physical Experimental Parameters

Experiment Parameter	Value
High Tracking Accuracy	No
Speed of RMT (cm/s)	15
Random movement of RMT (cm)	$a = 0$, $b = 0$
Number of loops	10

6.4.3 Validation of Physical Performance in Physical Human-Robot Collaboration

TABLE 6.4: Experimental Parameters Physical Performance Validation

Experiment Parameter	**Low Physical workload**	**Medium Physical workload**	**High Physical workload**
High Tracking Accuracy	No	No	No
Speed of RMT (cm/s)	5	10	15
Random movement of RMT (cm)	$a = 0$, $b = 0$	$a = 0$, $b = 0$	$a = 0$, $b = 0$
Number of Loops	10	10	10

The efficacy of the physical performance model has not previously been assessed in the context of pHRC. To address this gap, three experiments are designed, taking inspiration from Section 6.2.2 and referencing Table 6.4. These experiments mainly aim to determine whether variations in physical performance can induce differences in task performance.

A significant challenge lies in modulating the force applied. To overcome this, we adopted a three-tiered speed level for the RMT, the rationale being that a faster speed necessitates a higher force output. This is predicated on the direct proportional relationship established between the velocity of the end effector and the force applied, especially within the context of admittance control. By mirroring the pre-experiment presented in Section 6.4.1, we measure the maximum isometric force ($F_{max,iso}$) at second intervals as the human co-worker engages with the experiment.

The insights from this endeavour are vividly captured in Figure 6.8. This graphical representation underscores that the proposed physical performance model possesses a commendable capability to encapsulate human performance nuances. Figure 6.8 (a) dives into the details, presenting the force values harvested from the hand dynamometer over time in three different physical workload conditions. Under conditions of increased workload, both the culmination point of $F_{max,iso}$ and the gradient of the curve record the most subdued values. Figure 6.8 (c) corroborates this, demonstrating that the tracking error under high workload scenarios exhibits the most pronounced drop compared to the other two testing conditions. A deeper dive, propelled by the parameter identification method presented in Section 6.4.1, reveals the simulated physical performance. This simulation, which is based on specific physical parameters such as MVC, C_r, and Cf, finds a visual representation in Figure 6.8(b).

Analysing the data shown in Figure 6.8 leads to an interesting finding. It is evident that the variations in physical performance measurements, spread across the three different levels of physical workload, align in a specific order. Furthermore, the tracking errors among thesse scenarios manifest themselves differently. These observations show that the proposed physical performance model influences task performance when placed in the context of pHRC.

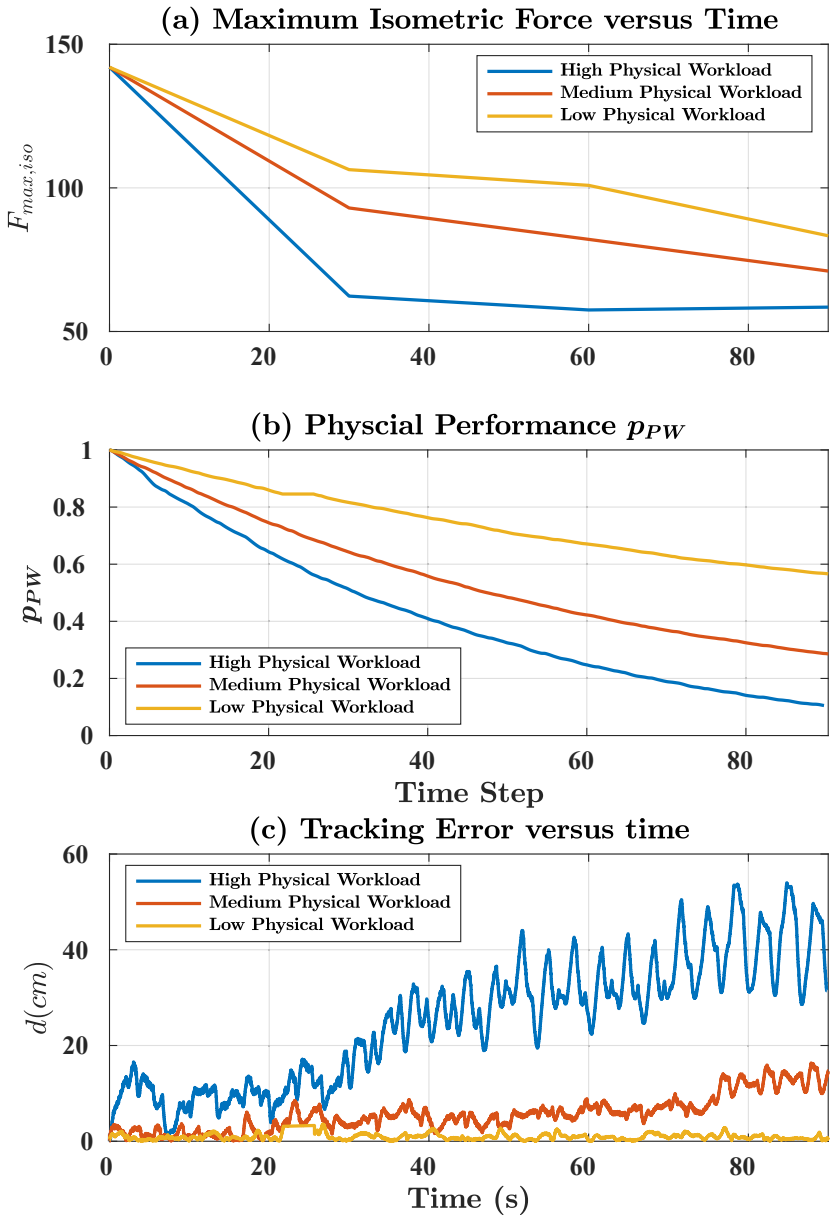


FIGURE 6.8: (a) Maximum isometric force $F_{max,iso}$ versus time under high, medium, and low physical workload (High, medium, and low speed of red moving target) (b) Physical Performance versus time (c) Tracking error versus time

6.4.4 Results and Discussion - Experiment 2 Physical Performance

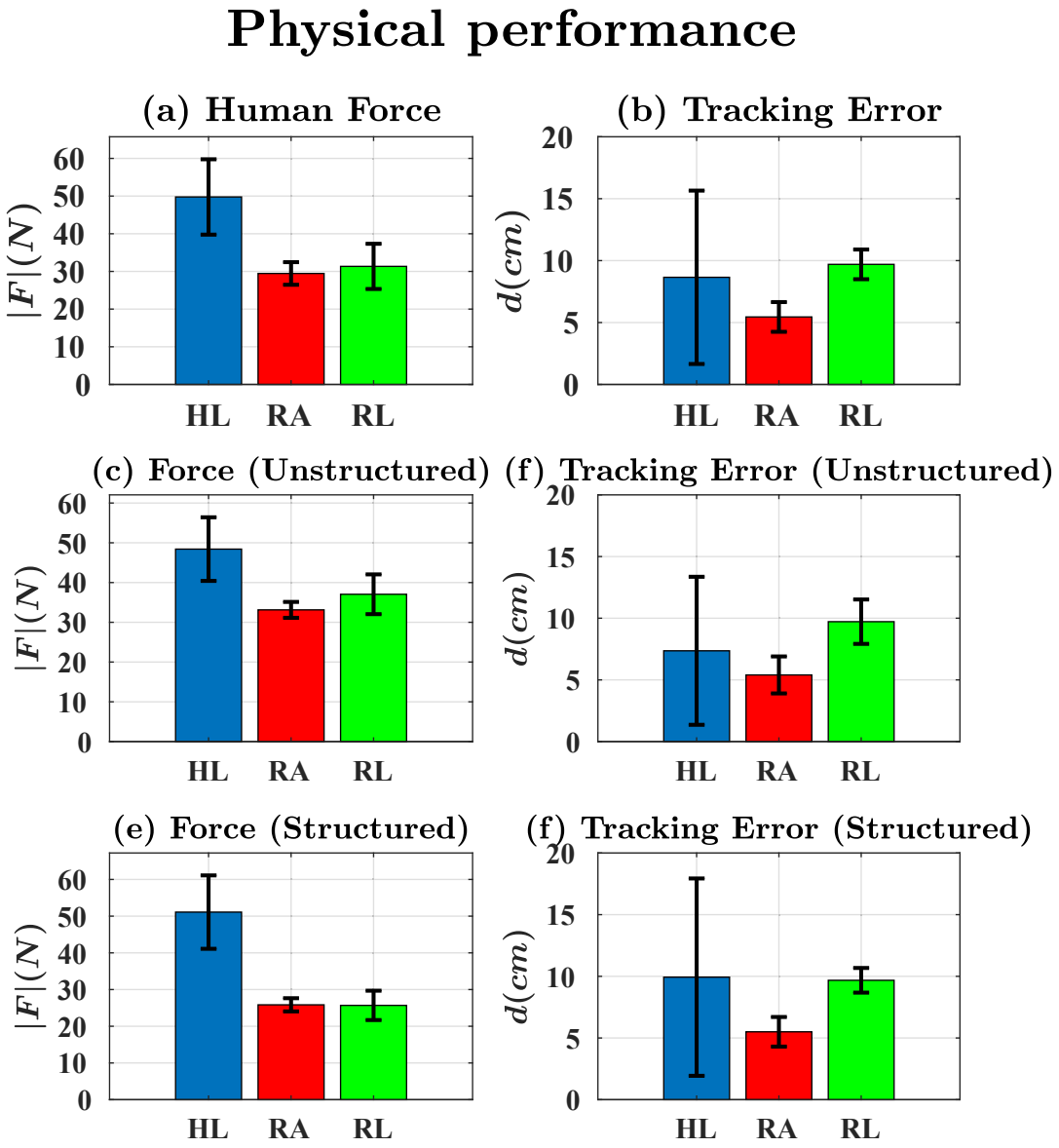


FIGURE 6.9: Evaluations of the three control strategies from the fifteen participants for physical performance in Experiment 2 in Section 6.3: (a) Human Force (Overall) (b) Tracking Error (Overall) (c) Human Force (Unstructured) (d) Tracking Error (Unstructured) (e) Human Force (Unstructured) (f) Tracking Error (Unstructured). Unstructured: First half of the trajectory sensor noise on the desired red moving target in Figure 6.1, i.e. from L1, L2 to L3. Structured: Second half of the trajectory without sensor noise, i.e. from L3, L4 to L1.

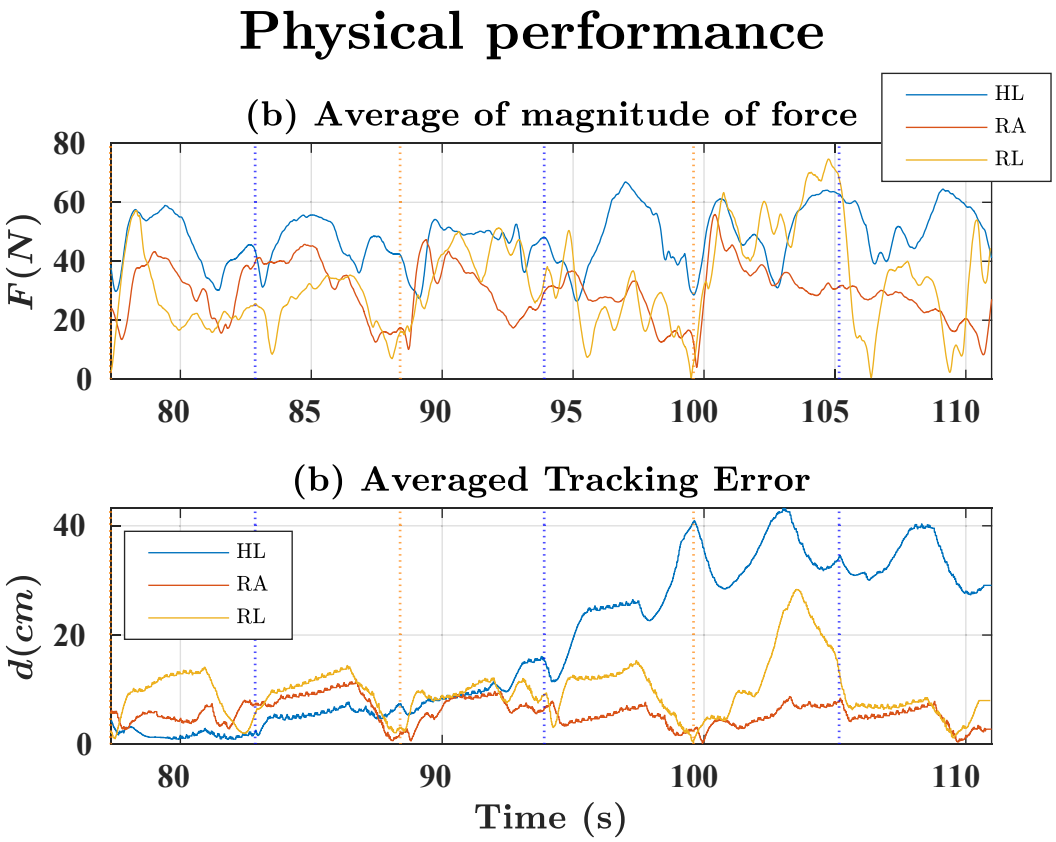


FIGURE 6.10: Evaluations of the three control strategies from the fifteen participants for physical performance in Experiment 2 in Section 6.4: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants over the experiment at each timestep. The vertical dotted lines are the timestep when transmitting between unstructured and structured environments (L1 and L3 in Figure 6.1). The red dotted vertical lines are the timestep when transmitting from structured to unstructured environment (L1 in Figure 6.1). And the blue dotted vertical lines are the timestep when transmitting from unstructured to structured environment (L3 in Figure 6.1). To clearly demonstrate the results, the last three loops of the results are presented in Figure 6.10. The full experimental duration results refer to Figure 6.11.

6.4.4.1 Average Magnitude of Human Force

1. **Over the Whole Trajectory:** Figure 6.9(a) shows that the RA control strategy can reduce the magnitude of human force by -43.30% and -47.10% compared to strategies of HL and RL, respectively.

2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.9(c) shows that RA reduces the magnitude of human force by -31.95% and -54.57% compared to HL and RL, respectively.

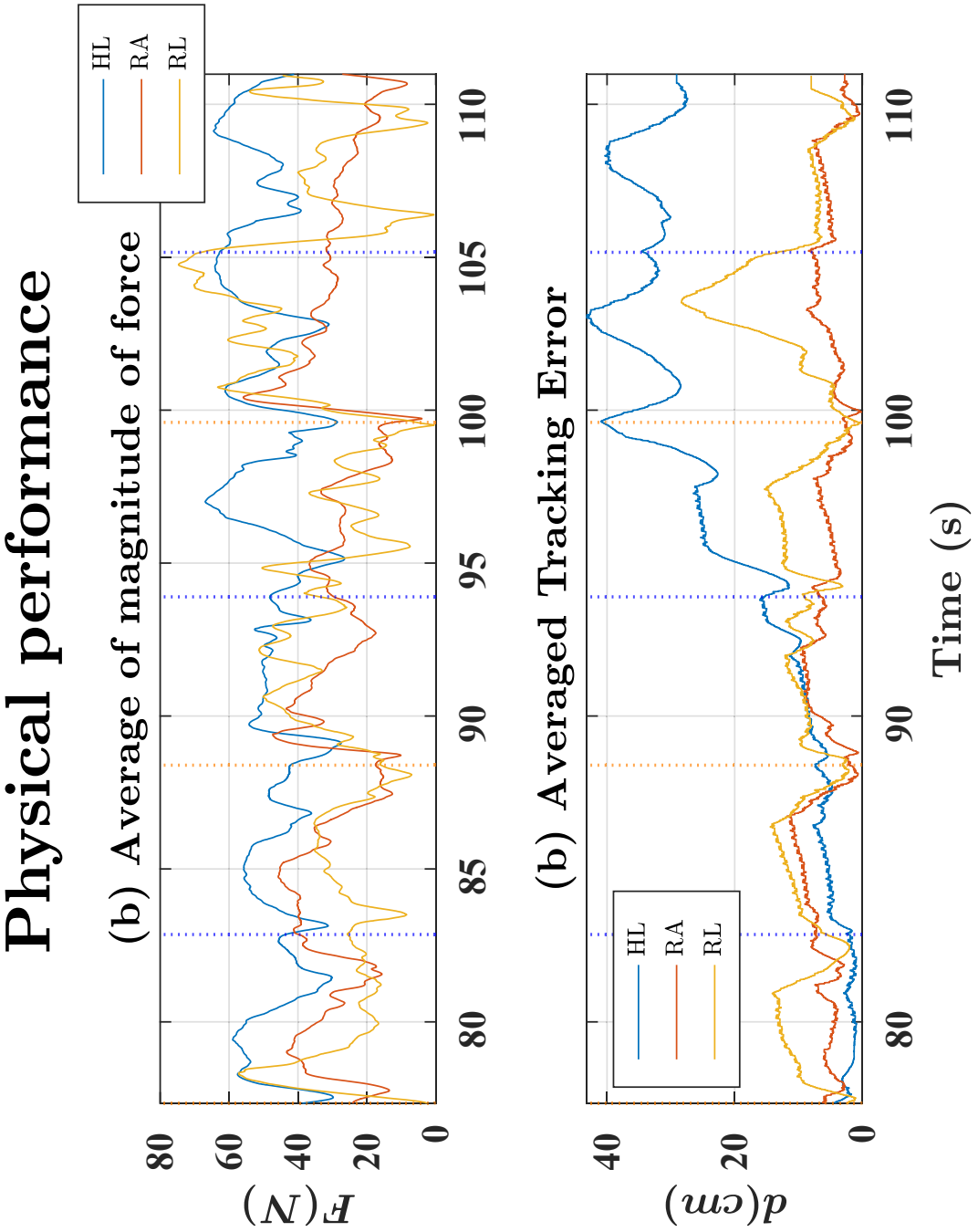


FIGURE 6.11: Evaluations of the three control strategies from the fifteen participants for Physical performance in Experiment 3 in Section 6.4: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants over the experiment at each timestep. Vertical dotted lines are the timestep when transmitting between an unstructured and a structured environment (L1 and L3 in Figure 6.1). The red dotted vertical lines are the timestep when transmitting from structured to unstructured environment (L1 in Figure 6.1).

3. **Structured (over the trajectory from L3, L4 to L1):** Figure 6.9(e) shows that the RA control strategy can reduce the magnitude of human force by -54.28% compared to the HL strategy, but increased by +30.65% compared to the RL strategy.

6.4.4.2 Average Tracking Error

1. **Over the Whole Trajectory:** Figure 6.9(b) shows that the RA control strategy can reduce the tracking error by -10.97% compared to the HL strategy, and by -13.11% compared to the RL strategy, respectively.

2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.9(d) shows that RA reduces the tracking error by -5. 95% compared to HL and increases by +28.56% compared to RL, respectively.

3. **Structured (over the trajectory from L3, L4 to L1)**: Figure 6.9(f) shows that RA reduces the tracking error by -16.07% compared to the HL strategy, but increases by +15. 43% compared to the RL strategy, respectively.

6.4.4.3 Discussion of human force and tracking error

In a comprehensive comparison, RA consistently outperforms both HL and RL in terms of human force and tracking errors, as illustrated in Figure 6.9(a) and (b). One noticeable observation during the experiment was the physical fatigue experienced by the participants when operating the robot. This led many to lean more towards autonomous control rather than actively tracking the RMT, which in turn resulted in pronounced tracking error escalations for RL. Such case can be distinctly ranked as $RL > HL > RA$ in Figure (b)(f)(e), which can be attributed to the sensor noise associated with autonomous control. Conversely, there is a marked reduction in human force, delineated as $HL >> RL > RA$ in Figure 6.9(a)(c)(e). This aligns with previous findings which suggest that human performance deteriorates when engagement levels are low [123]. A decrease in the force magnitude indicates increased reliance on autonomous control, but this often leads to compromised performance, evident from the rising tracking errors.

The pronounced variance (std) of HL in tracking error, as depicted in Figure 6.9(b)(d)(e), can be attributed to the varying strength levels among participants. At the end of the experiment, it was evident from Figure 6.10(b) that participants struggled to keep pace with the RMT, resulting in increased tracking errors. The exact moment of this surge largely depends on an individual's strength, endurance, and the rate at which fatigue sets in.

Regarding structured versus unstructured environments, the results concur with prior findings: autonomous control typically surpasses human control in structured settings, and the opposite holds true for unstructured scenarios. In contrast to previous experiments, the physical fatigue experienced by participants led to an increased reliance on autonomous control. This trend is evidenced by the insignificant standard deviation of RL in tracking error. Such behaviour suggests that human co-workers, perhaps due to diminished confidence in their abilities, placed greater trust in the robot to execute the task competently.

6.4.4.4 Failure Rate

Table 6.5 shows the failure rate of the experiment with thresholds of 30 cm. Due to the pronounced physical workload, particularly in HL, participants often find it challenging to complete the task independently. The need for assistance from autonomous control becomes evident by examining Figure 6.10. Consequently, the failure rate associated with HL emerges as the most significant.

TABLE 6.5: Physical experiment failure rate of HL, RA and RL based on the Evaluation Criteria 5 in Section 6.2.6

Threshold: 30cm	HL	RA	RL
Failure Rate %	62.5	0.00	0.00

6.4.4.5 Subjective - NASA-TLX

Figures 6.12 and 6.13 show the perceived workload of participants as captured by the NASA TLX questionnaire, using the RS and PC, respectively. A salient observation is the congruence between the findings of both methods. This significant alignment between the HL, RA and RL outcomes is corroborating with the insights from Chapter 5, suggesting that when differences are pronounced, the outcomes of RS and PC tend to converge.

When juxtaposed with the subjective results from the smoothness performance experiment 1, as seen in Figure 6.4 and Figure 6.5, it becomes evident that RA consistently outperforms both HL and RL in terms of interaction experience. The higher mental demands and levels of frustration for $RL > HL$ can be attributed to sensor noise inherent in autonomous control. These attributes are closely aligned with the unpredictability experienced while operating the robot. This resonates with previous findings suggesting that unpredictable movements increase cognitive dissonance in humans, and this increase has a strong association with increased mental demands and levels of frustration [85].

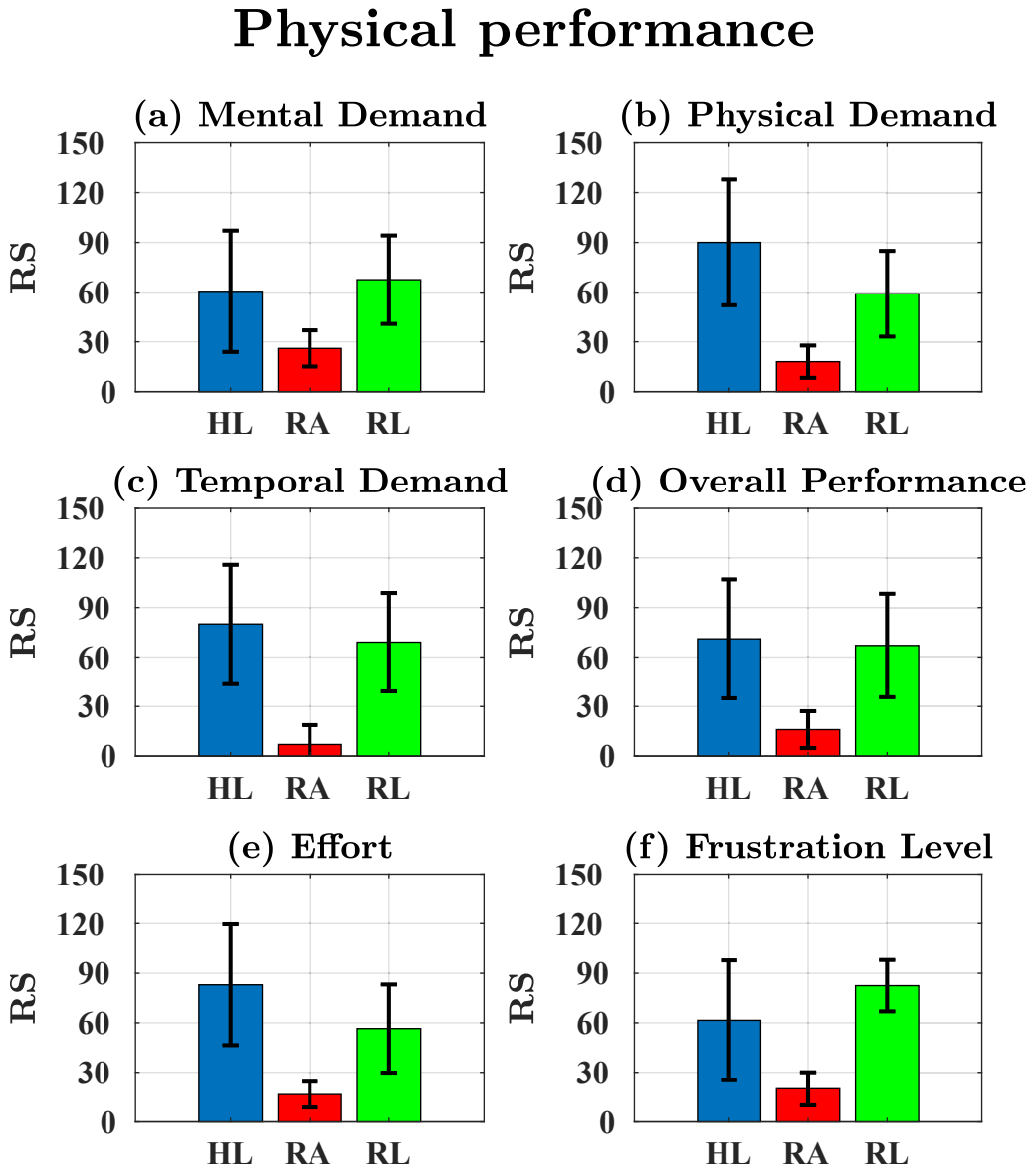


FIGURE 6.12: Normalized subjective results from NASA-TLX with rating scale method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants for smoothness experiment. 0 - good performance, 100 - bad performance.

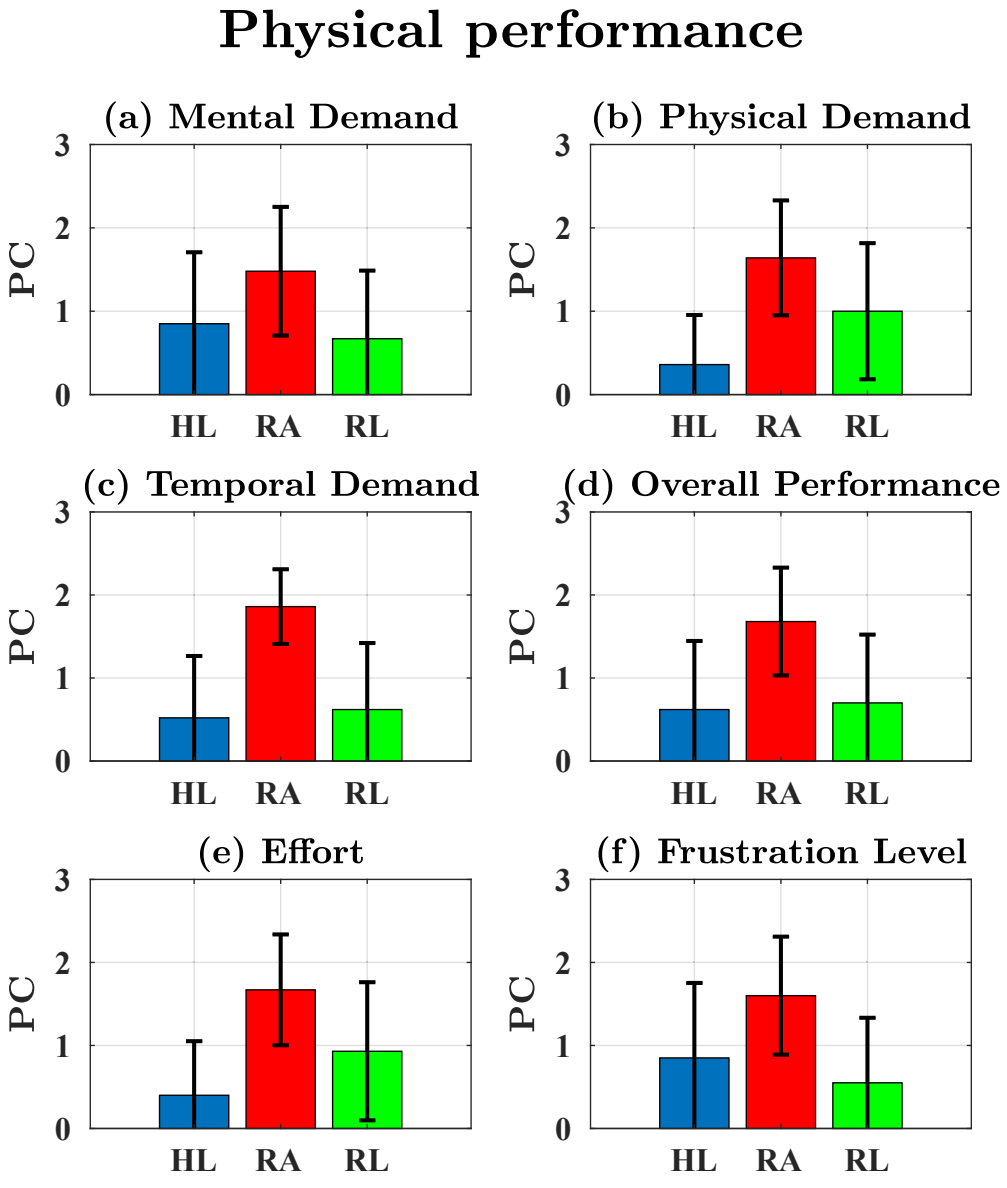


FIGURE 6.13: Subjective results from NASA-TLX with pairwise comparison method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants. 2 - good performance, 0 - bad performance.

Furthermore, factors such as physical demand, performance, temporal demand, and effort reveal a pattern of $HL > RL$, RL shows a better subjective impression compared to HL in those evaluation criteria, likely due to the pronounced physical fatigue encountered by participants. In particular, the standard deviation for HL and RL surpasses that of RA, underscoring the variability in individual strength levels that leads to different degrees of dependence on autonomous control. Participants with less strength tend to exhibit a marked preference for RL, valuing the consistent assistance despite occasional misguided control. On the contrary, participants endowed with higher strength display a predilection for HL, as they find misguided controls more irksome.

6.4.4.6 Conclusion - Physical Performance

An analysis using one-way ANOVA was performed, revealing statistically significant differences ($p < 0.05$) across all the evaluation metrics outlined in Section 6.2.3 between HL, RA, and RL. In summary, the data corroborates the efficacy of the proposed role arbitration method. It not only enhances the combined performance of human-robot teams but also alleviates the workload on human collaborators and augments subjective preference when integrated with physical performance metrics.

6.5 Experiment 3 - Cognitive Performance

6.5.1 Pre-experiment Parameter Identifcation

To accurately determine the task and individual-based cognitive performance parameters (a_r, p^+_{CP}, p^-_{CP}, and β) as required in Equations 3.15 and 3.16, it is crucial to account for the diversity in the proficiency levels of the participants and the varying degrees of difficulty of the task. The discrepancies between the expertise of the participants and the complexity of the tasks require this modelling process to ensure that cognitive performance is adequately represented for each individual and task scenario.

Therefore, an optimisation equation is formulated as shown in Equation 6.4 to optimise the cognitive performance parameters in Equation 3.15 and 3.16 by minimising the difference between task performance $T[n]$ and cognitive performance P_{CP} based on the hypothesis that the change in task performance during this task is assumed to be due to the change in cognitive performance only.

$$\min_{a_r, p^+_{CP}, p^-_{CP}, \beta} \quad \sum_{n=0}^{N} (T[n] - P_{CP}[n])^2 \tag{6.4}$$

Task performance $T[n]$ is task-dependent, defined as tracking error $d[n]$ for the current experiment design, which is the real-time difference between the current position of the end effector and the position of the RMT as shown in Equation 6.5. A normalisation function is defined based on the tracking error threshold d_{th} which can be obtained through expert trials (i.e., the maximum value of the expert trial).

$$T[n] = \frac{1}{W} \sum_{i=n-W}^{n} (1 - \frac{d[n]}{d_{th}}) \tag{6.5}$$

$d[n]$ is the tracking error at each timestep. W is the length of the time window. The moving average reduces the noise caused by human co-worker's random movement.

6.5.2 Experiment Design

In order to impose cognitive fatigue, an equal distributed noise is added to the position of the RMT (p_{RMT}) as shown in Equation 6.6. Visually from the participant, p_{RMT} changes randomly. The randomness of p_{RMT} is adjusted to increase cognitive workload by tuning the upper and lower limits of range a, b. In this experiment, $a = 0$ and $b = 8$ maximise the cognitive workload imposed on human participants based on the tests with experts who experience a high cognitive workload. In order to minimise the effect of physical fatigue imposed on the human co-worker, the red moving target is designed with low speed ($V_{RMT} = 2.5cm/s$). Details of the experimental parameters are shown in Table 6.6.

$$p^{H}_{RMT} = p_{RMT} + \mathcal{U}(a, b) \tag{6.6}$$

Table 6.6: Cognitive Experimental Parameters

Experiment Parameter	**Value**
High Tracking Accuracy Task Requirement	No
Speed of RMT (cm/s)	2.5
Random movement of RMT (cm)	$a = 0$, $b = 6$
Number of loops	2

Therefore, based on the pre-experiment parameter identification method shown in Section 6.5.1, the measured T and curve fitting P_{CP} based on the experiment design are shown in Figure 6.14.

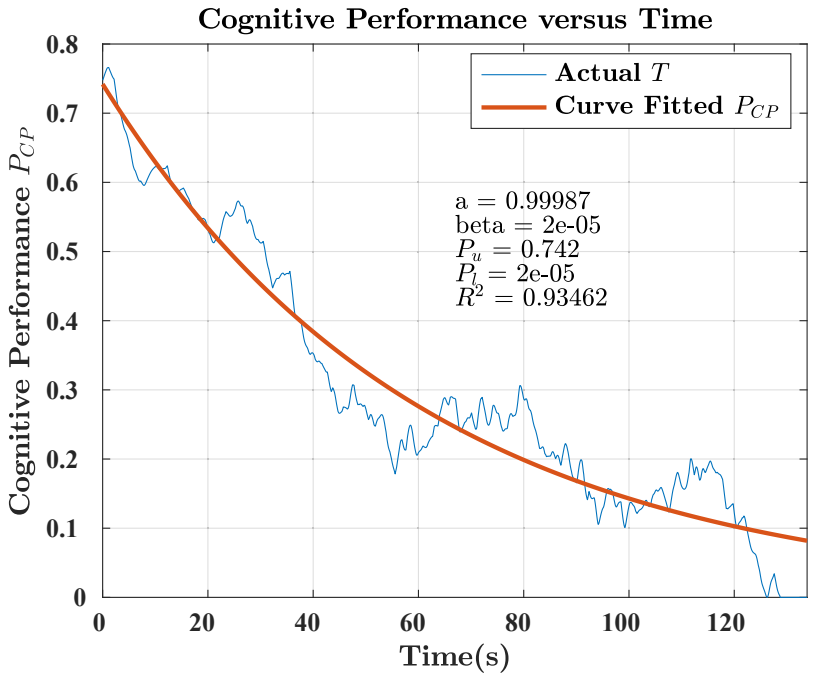


FIGURE 6.14: Measured task performance T (Equation 6.5) and curve fitting cognitive performance P_{CP} (Equation 6.4) versus time. This result is the average value of five trials from one participant. The identified cognitive performance parameters based on the curve fit result are $a_r = 0.9987, P_{CP}^{+} = 0.742, P_{CP}^{-} = 0, \beta = 0$, which corresponds to a high cognitive work load task that is shown in Figure 3.6. R^2 is the coefficient of determination which represents how well P_{CP} fit T.

TABLE 6.7: Experimental Parameters - Three level of difficulty experiments designed for validate Cognitive Performance through changing the randomness of movement of red moving target positions

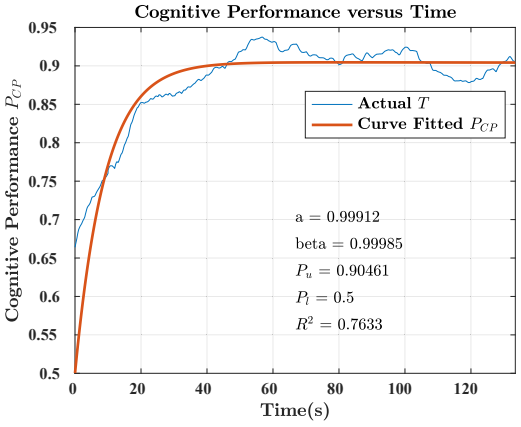
Experiment Parameter	**Basic Level**	**Intermediate Level**	**Advanced Level**
High Tracking Accuracy	No	No	No
Speed of RMT (cm/s)	2.5	2.5	2.5
Random movement of RMT (cm)	$a = 0$, $b = 2$	$a = 0$, $b = 4$	$a = 0$, $b = 6$
Number of Loops	2	2	2

6.5.3 Validation of Cognitive Performance

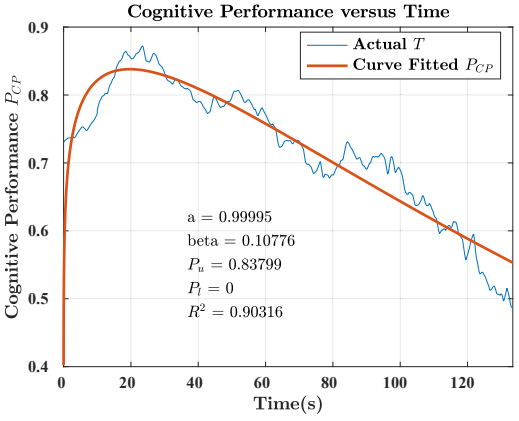
Since validation of the cognitive performance model in the context of pHRC has not previously been undertaken, this study aims to bridge that gap. Three experiments have been meticulously designed to assess whether the proposed cognitive performance model accurately gauges cognitive workload at varying levels of such workload. As delineated in Section 6.5, we adjusted the randomness of the position of the RMT, as illustrated in Table

6.7. This randomness is added to increase the cognitive difficulty of the task. The speed of RMT was minimised to limit any influence of physical workload on task performance.

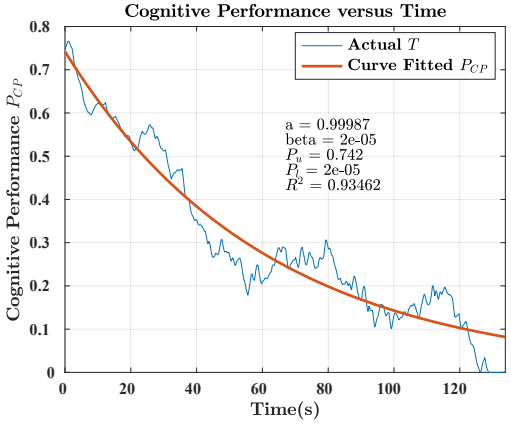
Figure 6.15d displays the measured maximum isometric force, which indicates a small decrease over time, which is 7% compared to 67% reduction in the maximum isometric force of the physical workload experiments. This indicates a stable physical workload throughout the experiment. Performance metrics for task T and curve adjustment P_{CP} at different levels of cognitive workload are depicted in Figures 6.15a, 6.15b, and 6.15c. These figures correspond to varying levels of task complexities.



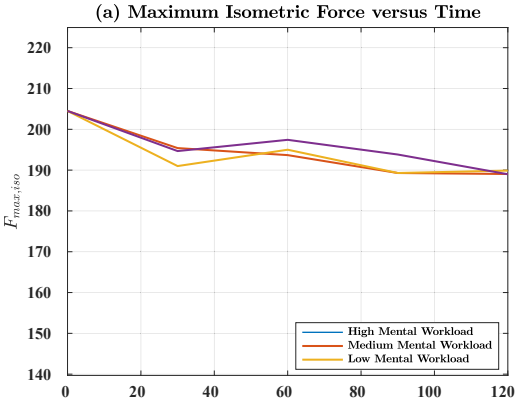
(a) Basic level



(b) Intermediate level



(c) Advanced level



(d) Measured Maximum Isometric Force under basic, intermediate and advanced levels task

FIGURE 6.15: Results of measured task performance T and curve fitted cognitive performance p_{CP} based on the optimization process in Section 6.5.1 under three different level of task difficulties (a) Basic Level (b) Intermediate Level (c) Advanced Level. The levels correspond to Table 6.7.

6.5.4 Results and Discussion - Experiment 3 Cognitive Performance

6.5.4.1 Average Magnitude of Human Force

1. **Over the Whole Trajectory:** Figure 6.16(a) shows that the RA control strategy can reduce the magnitude of human force by -32.79% and -33.55% compared to strategies of HL and RL, respectively.
2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.16(c) shows that RA reduces the magnitude of human force by -21.35% and -45.51% compared to HL and RL, respectively.
3. **Structured (over the trajectory from L3, L4 to L1):** Figure 6.16(e) shows that RA reduces the magnitude of human force by -43.68% and -14.46% compared to HL and RL, respectively.

6.5.4.2 Average Tracking Error

1. **Over the Whole Trajectory:** Figure 6.16(b) shows that the RA control strategy can reduce the tracking error by -2.06% compared to the HL strategy, and by -15.38% compared to the RL strategy.
2. **Unstructured (over the trajectory from L1, L2 to L3):** Figure 6.16(d) shows that the RA control strategy can reduce the tracking error by -9.11% compared to the RL strategy, but increased by +32.71% compared to the HL strategy.
3. **Structured (over the trajectory from L3, L4 to L1):** Figure 6.16(f) shows that the RA control strategy can reduce the tracking error by -12.50% compared to the HL strategy, but increased by +20.97% compared to the RL strategy.

6.5.4.3 Discussion of Human Force and Tracking Error

In a comprehensive assessment, the performance of RA consistently surpasses that of HL and RL, as demonstrated by the relationships $RL > HL > RA$ in Figure 6.16(a) for human force, and $RL \approx HL > RA$ in Figure 6.16(b) for tracking error. The performance trends of the three control strategies (HL, RA, and RL) mirror the results observed in the preceding two experiments. Specifically, in an unstructured environment, the relationship is denoted as $RA \approx HL < RL$ in Figure 6.16(c)(d) for both tracking error and human force. This can be attributed to the unpredictable movements generated by the robot, a consequence

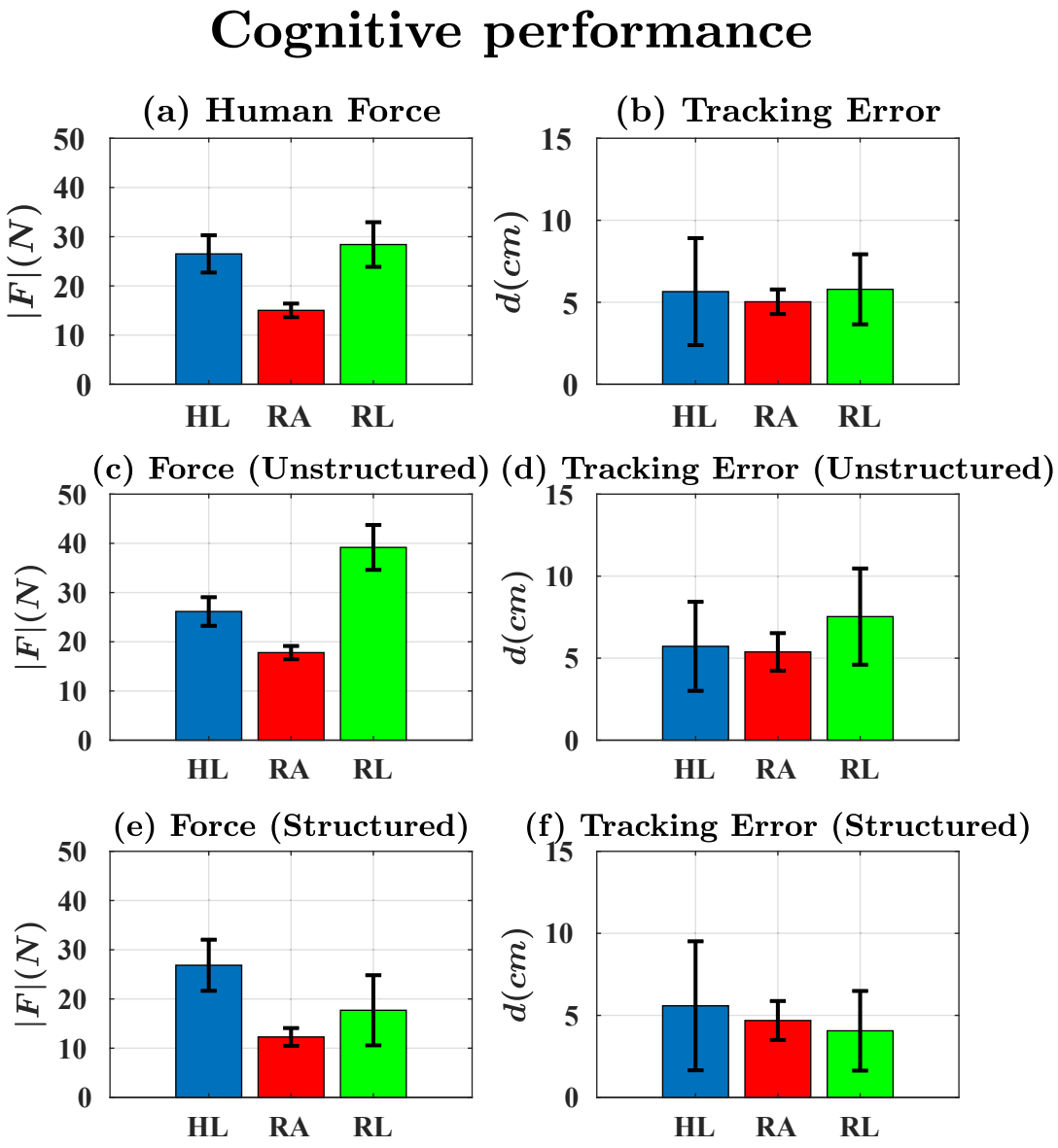


FIGURE 6.16: Evaluations of the three control strategies from the fifteen participants for cognitive performance in Experiment 3 in Section 6.5: (a) Human Force (Overall) (b) Tracking Error (Overall) (c) Human Force (Unstructured) (d) Tracking Error (Unstructured) (e) Human Force (Unstructured) (f) Tracking Error (Unstructured). Unstructured: The first half of the trajectory sensor noise on the desired red moving target in Figure 6.1, i.e. from L1, L2 to L3. Structured: Second half of the trajectory without sensor noise, i.e. from L3, L4 to L1.

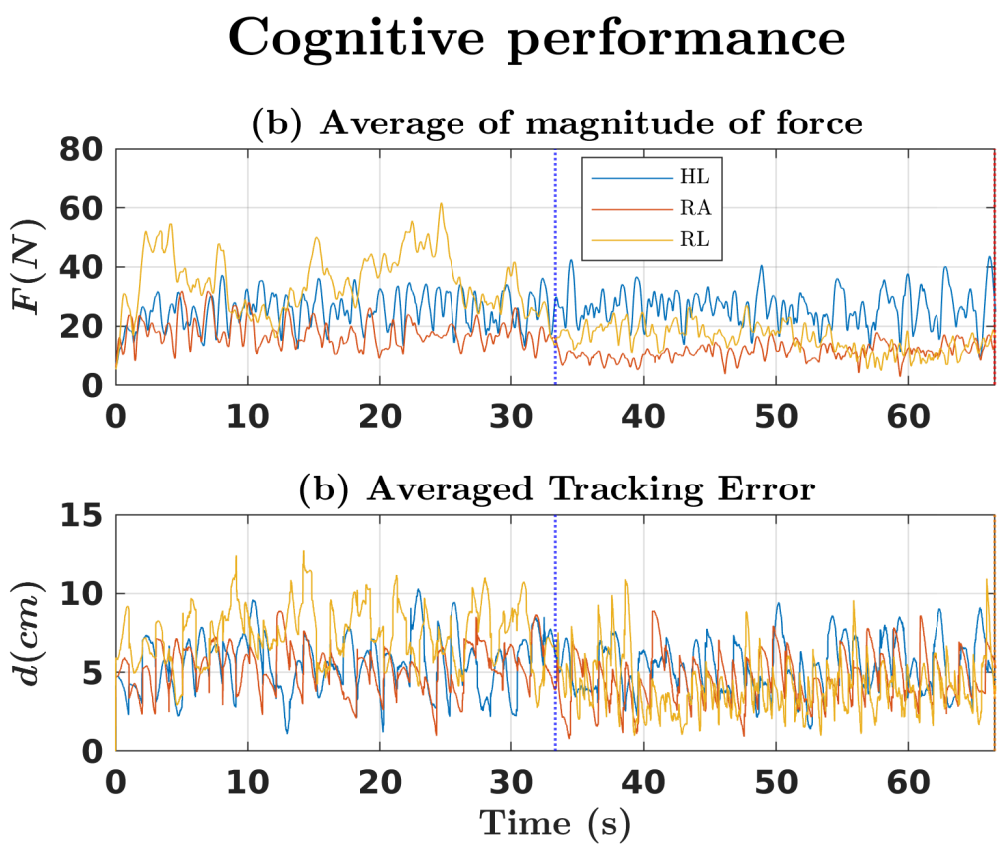


FIGURE 6.17: Evaluations of the three control strategies from the fifteen participants for Cognitive performance in Experiment 3 in Section 6.4: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants in the experiment at each timestep. The vertical dotted lines are the timestep when transmitting between unstructured and structured environments (L1 and L3 in Figure 6.1). The red dotted vertical lines are the timestep when transmitting from a structured to an unstructured environment (L1 in Figure 6.1). In order to demonstrate the results clearly, the first loop of results is presented in Figure 6.17. The full experimental duration results refer to Figure 6.18.

of sensor noise that affects the desired position of the RMT during autonomous control. However, in a structured setting, however, the sequence transitions to $RA \approx RL < HL$ (Figure 6.16(e)(f)), showcasing the inherent accuracy advantage autonomous control has over human intervention.

The substantial standard deviation (std) associated with HL for tracking error, as highlighted in Figure 6.16 (f), is due to the diverse levels of expertise of the participants regarding the designated task, the RI cohort (proficient group) is 2.04 cm, and the non-RI cohort (non-profieint group) is 9.12 cm. For instance, participants in the UTS RI institute cohort, who are presumably more adept, display reduced tracking errors, while external participants manifest a more pronounced error range. On the contrary, the tracking error standard deviations for RA and RL are considerably lower compared to HL. This can be

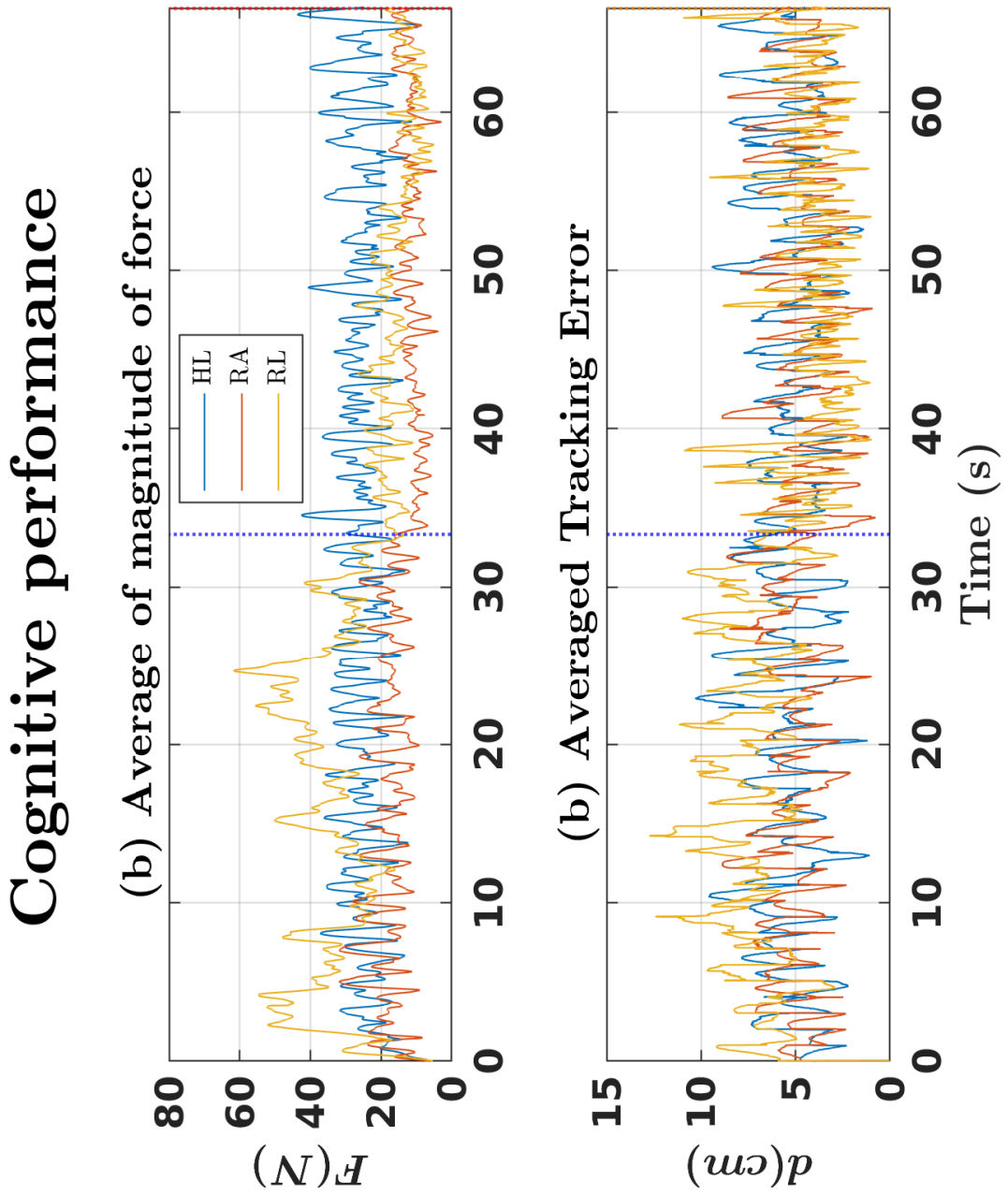


FIGURE 6.18: Evaluations of the three control strategies from the fifteen participants for Cognitive performance in Experiment 3 in Section 6.5: (a) Averaged Human Force (b) Averaged Tracking Error of the fifteen participants over the experiment at each timestep. The vertical dotted lines are the timestep when transmitting between unstructured and structured environments (L1 and L3 in Figure 6.1). The red dotted vertical lines are the timestep when transmitting from structured to unstructured environment (L1 in Figure 6.1).

explained by the influence of autonomous control dynamics within the role arbitration mechanism, rendering the robot's final pose more predictable than the purely manual input seen in HL. However, in an unstructured setting, RL shows a markedly higher std in human force (Figure 6.16(e)), a reflection of varying degrees of participant reliance on autonomous control. Notably, adept participants lean less on autonomous assistance, while their less experienced counterparts depend on it more, RI cohort (proficient group) is 27.43N, and the non-RI cohort (non-profieint group) is 9.84N.

6.5.4.4 Failure Rate

Table 6.8 shows the failure rates for cognitive performance when set against a 15 cm tracking error threshold. The findings delineate a trend as $RL > HL > RA$. The unpredictability in the movement of the RMT subjects the participants to an elevated cognitive workload. This surge in workload subsequently diminishes their focus on the task at hand, rendering them unable to accurately track the RMT's position under the HL condition. In the context of RL, the challenges are twofold: first, the unpredictability introduced by the robot due to misguided autonomous control due to the added sensor noise in RMT tracking, and second, the inherent unpredictability of RMT movements. Compound unpredictability makes it difficult for participants to control the robot effectively, leading to experiment failures.

TABLE 6.8: Cognitive experiment failure rate of HL, RA and RL based on the Evaluation Criteria 5 in Section 6.2.6

Threshold: 15 cm	HL	RA	RL
Failure Rate %	22.22	0.00	70.00

6.5.4.5 Subjective - NASA-TLX

Figures 6.19 and 6.20 show the results of the questionnaire based on both the rating scale (RS) and pairwise comparison (PC) methodologies, measuring participants' perceived workload through the NASA TLX framework. The alignment between the RS and PC results can be attributed to the marked disparities among RA, RL, and HL, a finding consistent with the conclusions drawn in Chapter 5. In particular, Figures 6.19 and 6.20 underscore RA's superior performance over RL and HL, the discrepancy being particularly pronounced for RL ($RL >> HL > RA$).

This observed trend can be traced back to pronounced unpredictability in both the robot movement and the designated task. The robot's unpredictable behaviour is a byproduct

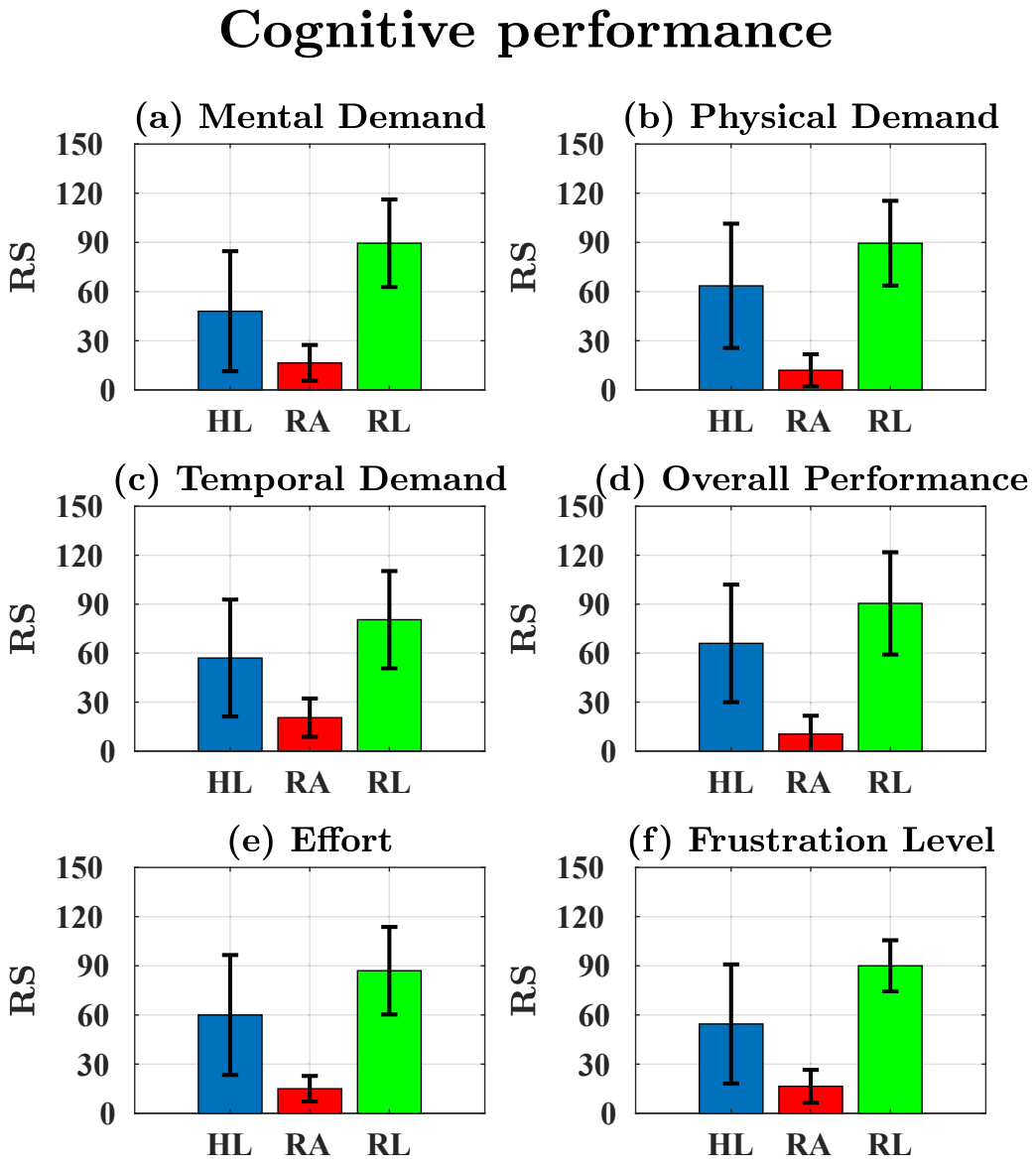


FIGURE 6.19: Normalized subjective results from NASA-TLX with rating scale method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants. 0 - good performance, 100 - bad performance.

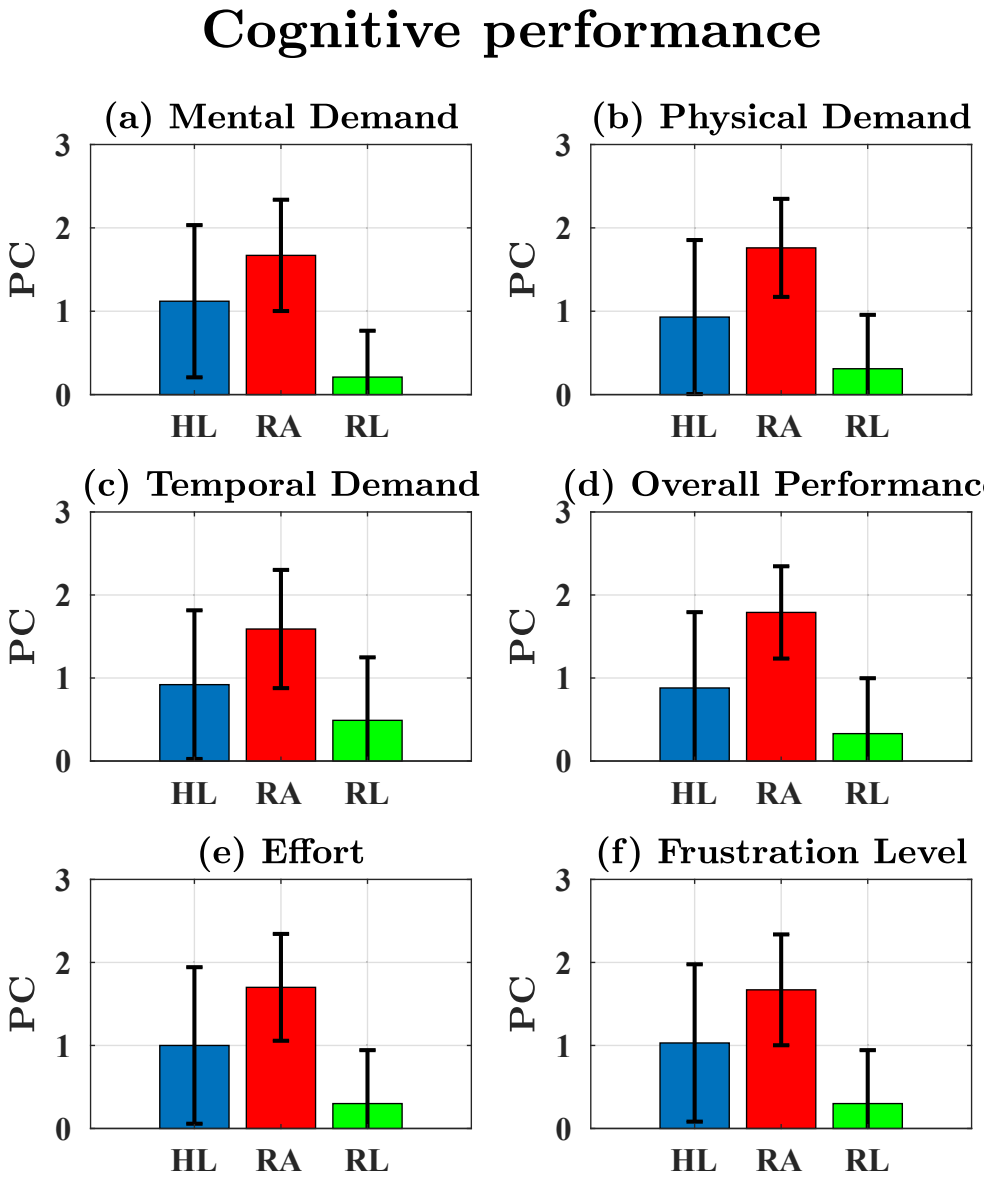


FIGURE 6.20: Subjective results from NASA-TLX with pairwise comparison method (Mental demand, physical demand, temporal demand, overall performance, effort, and frustration level) from the fifteen participants.

of sensor noise affecting the RMT's position, whereas the task's unpredictability arises from the random nature of RMT that was added to increase cognitive workload in this experiment. These dual unpredictability negatively affect the subjective experience of participants while navigating the robot within this cognitively demanding task. Consequently, the subjective impression of RL witnessed a substantial dip compared to previous experiments.

Upon a deeper analysis of individual subjective evaluation metrics, metrics such as mental workload, overall performance, and frustration levels surfaced as exceptionally elevated. This aligns with the inherent demands of the experiment's design, which was geared towards cognitive intensity. The intertwined unpredictability of both RMT and autonomous control spurred participants to yearn for greater control, manifesting in their application of increased force to reclaim command. In aggregate, the standard deviations (std) for HL and RL surpassed those of RA, stemming from variations in the participants' reliance on autonomous control and their task proficiency. Less adept participants exhibited greater sensitivity to misguided autonomous controls. For these individuals, any deviation in control resulted in a more drastic erosion of their subjective impression compared to their more proficient counterparts, the RI cohort (proficient group) is 57, and the non-RI cohort (non-profieint group) is 95 in rating scale comparison.

6.5.4.6 Conclusion - Cognitive Performance

A one-way ANOVA was performed, revealing statistically significant differences ($p < 0.05$) in all evaluation metrics outlined in Section 6.2.3 between HL, RA and RL. In summary, the data support the efficacy of the proposed role arbitration method. Not only does it amplify the synergistic performance of human-robot collaborations, but it also mitigates the workload on human counterparts and augments subjective preference, especially when integrated with metrics assessing cognitive performance.

6.6 Conclusion

In summary, this chapter details three experiments designed to validate the TSC-based role arbitration method introduced in Chapter 4. These experiments spanned performance measures that included smoothness, physical, and cognitive aspects. They were evaluated using both the traditional subjective evaluation method of rating scales and the pairwise comparison approach proposed in Chapter 5. Furthermore, the validity of cognitive and physical performance models in the context of pHRC was confirmed. The results demonstrate that the proposed models adeptly gauge the physical and cognitive workload of

human co-workers engaged in pHRC experiments. Based on the experimental results, the role arbitration method has proven its capability to enhance the synergistic performance of human-robot collaborations, alleviate human co-worker workload, and elevate subjective preference. These findings are in alignment with the conclusions drawn in Chapter 4.

6.6.1 Human Force and Tracking error

The RA control strategy consistently shows notable force reductions compared to HL in all performance measures, with reductions of $36.9\%, 40.8\%, 43.3\%$. When juxtaposed with RL, RA's force reductions are 8.16% and 5.19% in Exp-Smoothness and Exp-Cognitive, respectively. These disparities arise from participants' inclination towards autonomous control, influenced by their perceptions of superior tracking performance and the physical fatigue experienced during the task. Such a predisposition underscores participants' trust in autonomous control when they perceive their own capabilities as inadequate for the task. As a result, participants often minimise their force input, allowing the autonomous system to assume control.

In the domain of Exp-cognitive, RL's human force reduction stands at 47.1%, surpassing that of Exp-Smoothness and Exp-Physical. Such an outcome can be traced to the unpredictability inherent in both robot movement and the assigned task, with the former resulting from sensor noise impacting the RMT's position and the latter from RMT's randomness. This unpredictability prompts participants to exert greater force, indicating their preference for manual control over autonomous direction. The degree of trust that participants place in autonomous control directly influences the interaction force exerted on the robot, signaling their intent. Higher trust correlates with reduced force application. As an outcome, the RL tracking error in contexts of Exp-Smoothness and Exp-Physical is elevated due to heightened reliance on an autonomous control that can sometimes be misguided by sensor noise. On the contrary, in Exp-Cognitive, the tracking error decreases as participants become more engaged, reducing the dependence on autonomous control. This is consistent with research suggesting that human performance deteriorates during low-engagement activities [123].

The standard deviation for human force under RL is pronounced and is attributed to variations in individual trust in autonomous control and cognitive performance. Factors such as intrinsic trust dynamics, possibly influenced by psychological or cultural nuances, play a role. Furthermore, the self-assessed ability to complete the task emerges as a dominant factor. For example, participants familiar with robotics, such as those in the RI cohort, often exert more force due to their confidence. On the contrary, external participants, those

not in the RI cohort, generally rely more on autonomous control. However, these participants, unfamiliar with robotic systems, exhibit heightened sensitivity to unexpected or undesired robotic movements. Their subjective impression degrades more rapidly in cases of substantial discrepancies between the robot's desired and actual poses, in comparison to their proficient counterparts.

Consequently, future work should explore the integration of the proposed trust model within the broader landscape of human-robot interactions. Trust dynamics, which is inherently bidirectional in human-human interactions, extends similarly to human-robot interfaces. Investigating this bidirectional trust, which encompasses mutual trust and self-confidence of both humans and robots, offers a promising research trajectory. Within this framework, the development of a role arbitration method grounded in bidirectional trust dynamics could be particularly insightful.

6.6.2 Failure Rate

The failure rate serves as a critical benchmark for task completion. When a participant's tracking error exceeds a specified threshold, the task is deemed unsuccessful. With respect to this failure rate, RA consistently registers significantly lower values compared to RL and HL in all experiments. This can be attributed to the control method of RA being both predictable and intuitive for participants.

In the Exp-Physical, HL registers the highest failure rate, a consequence of participants being unable to fulfil task requirements, largely due to the pronounced physical fatigue they experience, especially towards the experiment's culmination. In the context of Exp-Smoothness and Exp-Cognitive, failure rates increase due to the misalignment between human control and autonomous control results in RL registers the highest failure rate. Participants often find themselves lacking the requisite strength or time to rectify the robot's undesired pose, induced by autonomous control, and align it with the desired position.

6.6.3 Subjective Impression - NASA-TLX

This chapter presents findings of a comprehensive evaluation using both the Rating Scale (RS) and Pairwise Comparison (PC) metrics. These results, as evidenced in Figures 6.4, 6.5, 6.12, 6.13, 6.19, and 6.20, align closely, confirming the observations of Chapter 5. Specifically, when control strategy differences are pronounced, RS and PC outputs align and both are able to effectively capture the difference in the participant's subjective impressions. Combined, the perceived workload of RA is reduced by 30.9%, 62.37%, and

41.77% relative to HL, and by 79.22%, 51.75%, and 80.25% compared to RL in smoothness, physical, and cognitive experiments.

6.6.3.1 Mental Demand and Frustration

Mental demands relate to objective pressures related to the task originating from the task, while frustration denotes the level of satisfaction and preference during the experimental activity [124]. In all three experiments, the trend is $RL >> HL > RA$. RL registers increased mental demands and frustration in comparison to HL and RA, primarily due to the unpredictable nature of autonomous control in unstructured environments. Both these aspects are intricately tied to predictability, with erratic robot movements escalating mental demands.

6.6.3.2 Physical Demand

Physical demands come from objective tasks [124]. The observed trend is $RL > HL > RA$ for both Exp-Smoothness and Exp-Cognitive. This outcome does not align with the objective measures of human force but does correlate in unstructured scenarios. One potential explanation is that participants intensify conflicting forces in unstructured situations, implying that mental demands can influence physical demands. In terms of Exp-Physical, HL predominantly dictates the trend, consistent with the attributes of physically demanding tasks.

6.6.3.3 Effort

Effort amalgamates the physical and cognitive demands [124]. In particular, cognitive elements play a pivotal role in motor tasks, particularly affecting movement stability [125]. Here, mental demands mirror the unpredictability intrinsic to autonomous control. Thus, effort encapsulates physical and mental pressures related to the task. Given the diverse nature of tasks, the relative weighting between these factors varies. For Exp-Cognitive and Exp-Smoothness, mental demand holds higher weight, resulting in $RL > HL > RA$. On the contrary, Exp-Physical places a premium on physical demands, leading to $HL > RL > RA$.

6.6.3.4 Temporal

Temporal demands peak in physically intensive tasks due to the rapid motion of RMT, positioning HL as the most demanding ($HL > RL > RA$). For Exp-Cognitive and

Exp-Smooth, RL leads due to the pronounced randomness of RMT. This requires the participants to maintain continuous focus on the task and adjust to the unpredictability introduced by autonomous control, resulting in the trend $RL > HL > RA$. Essentially, increased mental demands, spurred by unpredictability, amplify temporal demands.

6.6.3.5 Performance

Performance, as a metric, gauges the perceived quality of the execution of tasks by participants [124]. Generally, when interfacing with robots, human collaborators aim to curtail both their force input and tracking errors. In Exp-Physical, the emphasis is on the force magnitude due to the task's physically demanding nature, producing a trend of $HL >> RL > RA$, HL showing the worst performance in Exp-Physical. Conversely, for Exp-Smoothness and Exp-Cognitive, where physical demands are relatively subdued, tracking error becomes paramount. Erroneous autonomous controls, influenced by sensor noise, shift the robot away from the RMT's position, leading to the trend $RL > HL > RA$.

6.6.3.6 Discussion

The theme of unpredictability has emerged as a decisive factor that influences subjective perception in this study. All dimensions of the NASA-TLX evaluation are influenced by unpredictability to varying extents. In the realm of pHRC, predictability can be articulated as the deviation between the desired and actual postures of the robots. According to Maurice et al., while manipulating intricate objects, people strive to improve task predictability [126]. This tendency to forecast the results of the tasks is derived from accumulated experiences, reflecting the proficiency or strength of the participants [127]. Aldini's studies, using physiological markers such as EEG, revealed that variations in cognitive conflict arise from disparities between the intended and actual pose of the robot in pHRC [83, 85].

Collectively, the RA strategy has demonstrated superior performance over HL and RL. In Exp-smoothness and Exp-cognitive, the trend of $RL > HL > RA$ emerges, mainly due to the elevated importance of mental demands influencing subjective perception. This is exacerbated by the unpredictability and errors introduced by autonomous control. On the contrary, during Exp-Physical, the emphasis shifts to physical demand due to the intensity of the tasks, resulting in the trend $HL > RL > RA$. Feedback from participants indicated that if autonomous control could consistently align with human intentions, leading to precise tracking performance, subjective experience would be positively enhanced. This alignment is particularly pronounced in structured environments, although its influence diminishes in unstructured settings. An exception was noted in physical tasks where, despite

deviations, the autonomous control remains within the range of the RMT, corroborated by the RL's low failure rate.

The harmony between human control and autonomous actions, striving towards a unified goal, stands out as a pivotal determinant of subjective perception in role arbitration research. This is owing to the integrative and physically tethered nature of pHRC. A misalignment could potentially lead to the under-utilisation or misuse of robots. Individual variations, as indicated by the standard deviations of HL, RA, and RL, further highlight the complexities. Standard deviations of HL and RL are notably higher than those of RA, and HL is influenced by individual parameters such as strength or proficiency. RL's variability, on the other hand, stems from differing degrees of reliance on autonomous control. Less proficient participants show a rapid decrease in performance and those with reduced physical strength or expertise are more susceptible to misguidance from autonomous control. On the contrary, experts or those with higher strength levels have a more predictable understanding of robot movements and tasks, making them less influenced by these factors. This all circles back to the overarching theme of predictability, which can be modulated by the robot, the task, proficiency levels, and the alignment between the task's objective and the robot's movements.

Chapter 7

Conclusion and Future Work

Due to increasing concerns about ageing issues and work-related muscle disorders, physical human-robot collaboration becomes an effective solution to address these concerns. To emulate human-human interaction in human-robot interaction, one of the important human-like cognitive processes is developed: trust and self-confidence. According to human factor research, an agent tends to rely on its collaborator when trust in their collaborator is greater than self-confidence, and vice versa [34]. Therefore, a role arbitration method based on robot trust in human co-workers and robot self-confidence is proposed. In addition, subjective impression is critical in pHRC as a result of the constant physical coupling between humans and robots. The long duration of negative feelings on the robot may cause disuse and degraded human-robot joint performance. A pairwise comparison questionnaire evaluation method is proposed to capture subjective impression with subtle differences in pHRC.

Based on the aforementioned motivation, a research question is proposed:
Q: How can trust and self-confidence be used to create effective arbitration in pHRC?

To elucidate the main research question, the following sub-questions have been formulated:
Q1: What are the effective metrics for measuring robot trust in human collaborators within pHRC?
Q2: How can robot self-confidence in pHRC be gauged?
Q3: How might we arbitrate roles based on a robot's trust in human co-workers and its self-confidence?
Q4: Which techniques are most effective and efficient to assess human subjectivity in pHRC?

Chapter 3 addresses sub-research question Q1: A computational model of robot trust in a human co-worker was proposed. This model takes into account many factors in the physical human-robot collaboration, including robot safety, robot singularity, smoothness of robot motion, and human physical and cognitive performance. Three experiments were conducted to verify all the factors in the model. Experiment 1 demonstrates variability in the rate of change in singularity performance as the system transitions into and out of singular configurations. Notably, alterations in the system's direction or velocity are correlated with changes in performance smoothness. In Experiment 2, findings indicate a positive correlation between the magnitude of applied force and enhancements in physical performance. Similarly, an extension in the duration of the experiment is associated with improvements in cognitive performance metrics. Experiment 3 reveals fluctuations in safety performance metrics contingent upon the proximity to surrounding obstacles, illustrating a dynamic response to the spatial relationship with nearby impediments.

Chapter 4 addresses sub-research questions Q2 and Q3: This thesis proposed and experimentally verified a TSC-based role arbitration method with a human subjects experiment that included a robot trust in human model, robot self-confidence, and a role arbitration method using the aforementioned two trust models. The proposed role arbitration method has been shown to achieve superior combined human-robot performance, reduce the workload of human coworkers, and improve subjective preference. The results show that the RA control strategy can reduce the magnitude of human force by 44.9% and 40.6% compared to the HL and RL strategies, respectively. Furthermore, RA reduces the tracking error by 18.4% compared to HL and increases by 10.3% compared to RL. However, RL has the highest failure rate, among the three control strategies, at 46.5%. In terms of subjective evaluation, RA reduces 67.3% and 79.8% compared to HL and RL, respectively.

Chapter 5 addresses sub-research questions Q4: A novel comparison between two subjective impression evaluation methods is conducted. The classic NASA-TLX questionnaire was used as the template to create the Rating Scale (RS) and Pairwise Comparison (PC) questionnaires, respectively. Through two experimental procedures of different lengths, the effectiveness of RS and PC was statistically compared. The experimental results show statistically significant results for both RS and PC when the difference between the comparing groups is large ($p < 0.05$ for both RS and PC). However, when the true difference is small, PC is more robust and accurate than RS regardless of the experimental procedure ($p > 0.05$ for RS and $p < 0.05$ for PC). Furthermore, compared to RS, PC also reduced the duration of the experiment by 44%, resulting in a more pleasant experimental experience and lower cognitive workloads.

Chapter 6 addresses the overarching Research Question experimentally: By combining trust models and role arbitration, as well as the pairwise comparison method in Chapters

3, 4, and 5. A comprehensive experiment is conducted that incorporates smoothness, physical and cognitive performance. Physical and cognitive performance has been validated in the context of pHRC. Similarly to the results in Chapter 4, the proposed role arbitration method has been shown to achieve superior combined human-robot performance, reduce the workload of human coworkers, and improve subjective preference. The difference between the proposed role arbitration method and two other fixed arbitration methods (human-leading and robot-leading) is significant. Both the rating scale and pairwise method could capture the subjective impression of the human coworker in pHRC, which is compliant with the finding in Chapter 5. The performance of RA is increased by 36.90%, 32.31%, and 10.97% relative to HL, and by 43.78%, 38.65%, and 13.11% compared to RL in smoothness, physical, and cognitive experiments, respectively. The physical workload of RA is reduced by 36.99%, 40.8%, and 43.3% relative to HL, and by 8.16%, 5.99%, and 47.1% compared to RL in smoothness, physical, and cognitive experiments. The perceived workload of RA is reduced by 30.9%, 62.37%, and 41.77% relative to HL, and by 79.22%, 51.75%, and 80.25% compared to RL in smoothness, physical, and cognitive experiments.

The outcomes of this research demonstrate the importance of appropriate computational trust model in achieving effective pHRC. Whether the trust model method could also be effective in other fields of HRI can be an interesting research direction to be investigated. For example, in telemanipulated surgical robotics, the trust model with its Safety performance metrics may serve to assess discrepancies between the configurations of robots and the human body, ensuring that a surgical robot's manipulator maintains a safe distance from the operation site. Singular performance evaluations address the inherent characteristic of manipulators to encounter singular configurations, highlighting the importance of understanding these conditions for effective operation. Smoothness performance, meanwhile, is pertinent to assessing the control accuracy of human co-workers, a critical aspect in ensuring seamless human-robot collaboration. Cognitive performance gains prominence, particularly in medical applications, where the demand for sustained attention is high. The extended duration of surgical tasks can impose significant physical strain on surgeons, especially when controlling the robot, thereby underlining the importance of monitoring physical performance to mitigate fatigue and enhance efficacy.

The trust model may also have additional applications for social robotics, where the intricacies of social HRI pivot significantly on cognitive performance, given its impact on the quality of interaction. The integration of cognitive performance assessments with additional factors such as robot appearance and physiological signals, like EEG data, presents a fertile avenue for research. This opens up possibilities for leveraging computational models of robot trust in humans across a wider array of HRI applications beyond physical collaboration, suggesting a holistic approach to understanding and enhancing human-robot interactions.

In regards to role arbitration, the concepts of robot trust and self-confidence extend their applicability to various fields within HRI. For example, in surgical telemanipulation, advancements in autonomous control technologies, fueled by significant progress in surgical vision pattern recognition and three-dimensional reconstruction, enable robots to more accurately navigate to targeted areas of human tissue. Despite these technological advancements, the inherent complexity of surgical patterns and the dynamic nature of human tissue deformation present challenges. These challenges manifest as errors in the geometric and semantic mapping of the surgical environment, subsequently impacting the precision of autonomous control.

Future work includes how the proposed trust model can fit into a larger context of trust in human-robot interaction, such as social interaction. Due to the trust being bidirectional in human-human interaction, combining the proposed robot's trust and self-confidence model with the human's trust and self-confidence model to become a more comprehensive trust dynamic in the human-robot paradigm can be another interesting research topic to be investigated. A new trust-based role arbitration paradigm can be modelled based on this bidirectional trust dynamics. More specifically, the determination of robotic intention is influenced by the discrepancy between the robot's trust in humans and its own self-confidence. Conversely, human intention is shaped by the variance between human trust in the robot and the individual's self-confidence. Therefore, improving the control mechanism of the robot, centered on understanding and optimising the dynamics of bidirectional interaction between humans and robots, presents an intriguing area of research. Such an exploration could elucidate how these interdependent factors - trust and confidence on both sides - dynamically influence the control allocation process, offering insights into the development of more responsive and intuitive human-robot collaboration systems.

The findings of this thesis have significant implications for the design and implementation of effective and enjoyable human-robot interactions, especially in scenarios that require physical contact and cooperation. The thesis also opens up new avenues for future research, such as extending the trust model to other domains of HRI and exploring the bidirectional dynamics of trust and self-confidence between humans and robots. This thesis demonstrates the potential of using trust and self-confidence as key factors for creating harmonious and productive human-robot teams.

Bibliography

[1] Marc G. Carmichael, Stefano Aldini, Richardo Khonasty, Antony Tran, Christian Reeks, Dikai Liu, Kenneth J. Waldron, and Gamini Dissanayake. The anbot: An intelligent robotic co-worker for industrial abrasive blasting. *IEEE International Conference on Intelligent Robots and Systems*, pages 8026–8033, 2019. ISSN 21530866. doi: 10.1109/IROS40897.2019.8967993.

[2] Marc G. Carmichael, Dikai Liu, and Kenneth J. Waldron. A framework for singularity-robust manipulator control during physical human-robot interaction. *International Journal of Robotics Research*, 2017. ISSN 17413176. doi: 10.1177/0278364917698748.

[3] H. Krebs, J. Palazzolo, Laura Dipietro, M. Ferraro, J. Krol, K. Rannekleiv, Bruce Volpe, and Neville Hogan. Rehabilitation robotics: Performance-based progressive robot-assisted therapy. *Autonomous Robots*, 15:7–20, 07 2003. doi: 10.1023/A:1024494031121.

[4] Susanna Freivogel, Jan Mehrholz, Tanya Husak-Sotomayor, and Dieter Schmalohr. Gait training with the newly developed 'lokohelp'-system is feasible for non-ambulatory patients after stroke, spinal cord and brain injury. a feasibility study. *Brain Injury*, 22(7-8):625–632, 2008.

[5] John Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992. ISSN 13665847. doi: 10.1080/00140139208967392.

[6] Anqi Xu and Gregory Dudek. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. *ACM/IEEE International Conference on Human-Robot Interaction*, 2015-March:221–228, 2015. ISSN 21672148. doi: 10.1145/2696454.2696492.

[7] S. M. Mizanoor Rahman, Y. Wang, I. D. Walker, L. Mears, R. Pak, and S. Remy. Trust-based compliant robot-human handovers of payloads in collaborative assembly in flexible manufacturing. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 355–360, 2016. doi: 10.1109/COASE.2016.7743428.

[8] Antony Tran, Dikai Liu, Ravindra Ranasinghe, and Marc Carmichael. A method for quantifying a robot's confidence in its human co-worker in human-robot cooperative grit-blasting. *50th International Symposium on Robotics, ISR 2018*, pages 474–481, 2018.

[9] J. Edward Colgate, Witaya Wannasuphoprasit, and Michael A. Peshkin. Cobots: robots for collaboration with human operators. *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC*, 58:433–439, 1996.

[10] Scott Green, Mark Billinghurst, Xiaoqi Chen, and James Chase. Human robot collaboration: An augmented reality approach a literature review and analysis. 01 2007. doi: 10.1115/DETC2007-34227.

[11] Zhijiang Du, Yunlei Liang, Zhiyuan Yan, Lining Sun, and Wei Chen. Human-robot interaction control of a haptic master manipulator used in laparoscopic minimally invasive surgical robot system. *Mechanism and Machine Theory*, 156:104132, 2021. ISSN 0094-114X. doi: https://doi.org/10.1016/j.mechmachtheory.2020.104132.

[12] L. Masia, H. I. Krebs, P. Cappa, and N. Hogan. Whole-arm rehabilitation following stroke: Hand module. *Proceedings of the First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006, BioRob 2006*, 2006: 1085–1089, 2006. doi: 10.1109/BIOROB.2006.1639236.

[13] H. I. Krebs, J. J. Palazzolo, L. Dipietro, M. Ferraro, J. Krol, K. Rannekleiv, B. T. Volpe, and N. Hogan. Rehabilitation robotics: Performance-based progressive robot-assisted therapy. *Autonomous Robots*, 15:7–20, 2003. ISSN 09295593. doi: 10.1023/A:1024494031121.

[14] Haoyong Yu, Sunan Huang, Gong Chen, Yongping Pan, and Zhao Guo. Human-robot interaction control of rehabilitation robots with series elastic actuators. *IEEE Transactions on Robotics*, 31:1089–1100, 2015. ISSN 15523098. doi: 10.1109/TRO.2015.2457314.

[15] Ernesto Gambao, Miguel Hernando, and Dragoljub Surdilovic. A new generation of collaborative robots for material handling. *2012 Proceedings of the 29th International*

Symposium of Automation and Robotics in Construction, ISARC 2012, 2012. ISSN 1569-1101. doi: 10.4017/gt.2012.11.02.362.00.

[16] Australian Government. 2023 intergenerational report. Technical report, Treasury, 2023. URL `https://treasury.gov.au/publication/2023-intergenerational-report`.

[17] Department of Economic United Nations and Population Division Social Affairs. World population prospects: Probabilistic population projections, percentage of population aged 65 or over, 2023. URL `https://population.un.org/wpp/Graphs/Probabilistic/PopPerc/65plus/900`. Accessed: 25 09 2023.

[18] J S Boschman, H F van der Molen, J K Sluiter, et al. Musculoskeletal disorders among construction workers: a one-year follow-up study. *BMC Musculoskeletal Disorders*, 13:196, 2012. doi: 10.1186/1471-2474-13-196.

[19] Laura Punnett and David H. Wegman. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology*, 14(1):13–23, 2004. ISSN 1050-6411. doi: https://doi.org/10.1016/j.jelekin.2003.09.015. State of the art research perspectives on muscoskeletal disorder causation and control.

[20] Luis Ortiz-Hernández, Silvia Tamez-González, Susana Martınez-Alcántara, and Ignacio Méndez-Ramırez. Computer use increases the risk of musculoskeletal disorders among newspaper office workers. *Archives of Medical Research*, 34(4):331–342, 2003. ISSN 0188-4409. doi: https://doi.org/10.1016/S0188-4409(03)00053-5.

[21] Freek Lötters, Willem-Jan Meerding, and Alex Burdorf. Reduced productivity after sickness absence due to musculoskeletal disorders and its relation to health outcomes. *Scandinavian journal of work, environment & health*, pages 367–374, 2005.

[22] Tomohisa Nagata, Koji Mori, Makoto Ohtani, Masako Nagata, Shigeyuki Kajiki, Yoshihisa Fujino, Shinya Matsuda, and Ronald Loeppke. Total health-related costs due to absenteeism, presenteeism, and medical and pharmaceutical expenses in japanese employers. *Journal of Occupational and Environmental Medicine*, 60:1, 02 2018. doi: 10.1097/JOM.0000000000001291.

[23] K M Oude Hengel, B M Blatter, G A Geuskens, L L Koppes, and P M Bongers. Factors associated with the ability and willingness to continue working until the age of 65 in construction workers. *International Archives of Occupational and Environmental Health*, 85(7):783–790, Oct 2012. doi: 10.1007/s00420-011-0719-3. Epub 2011 Nov 23.

[24] W.J. Meerding, W. IJzelenberg, M.A. Koopmanschap, J.L. Severens, and A. Burdorf. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *Journal of Clinical Epidemiology*, 58(5):517–523, 2005. ISSN 0895-4356. doi: https://doi.org/10.1016/j.jclinepi.2004.06.016.

[25] Ravindra Ranasinghe, Lakshitha Dantanarayana, Antony Tran, Stefan Lie, Michael Behrens, and LiYang Liu. Smart hoist: An assistive robot to aid carers. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, pages 1285–1291, 2014. doi: 10.1109/ICARCV.2014.7064501.

[26] Cynthia Breazeal, Atsuo Takanishi, and Tetsunori Kobayashi. *Social Robots that Interact with People*, pages 1349–1369. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-30301-5. doi: 10.1007/978-3-540-30301-$5_5$9.

[27] John D. Lee and Katrina A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004. doi: 10.1518/hfes.46.1.50_30392. PMID: 15151155.

[28] Michael Lewis, Katia Sycara, and Phillip Walker. *The Role of Trust in Human-Robot Interaction*, pages 135–159. Springer International Publishing, Cham, 2018. ISBN 978-3-319-64816-3. doi: 10.1007/978-3-319-64816-3_8.

[29] Bonnie M. Muir. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5):527–539, 1987. ISSN 0020-7373. doi: https://doi.org/10.1016/S0020-7373(87)80013-5.

[30] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O'Malley. A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction, 2018. ISSN 00036900.

[31] Stefano Aldini. Prediction-error negativity in physical human-robot collaboration. 2021.

[32] Luka Peternel, Nikos Tsagarakis, Darwin Caldwell, and Arash Ajoudani. Adaptation of robot physical behaviour to human fatigue in human-robot co-manipulation. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 489–494, 2016. doi: 10.1109/HUMANOIDS.2016.7803320.

[33] Marc G. Carmichael and Dikai Liu. Estimating physical assistance need using a musculoskeletal model. *IEEE Transactions on Biomedical Engineering*, 60(7):1912–1919, 2013. doi: 10.1109/TBME.2013.2244889.

[34] Stephan Lewandowsky, Michael Mundy, and Gerard Tan. The dynamics of trust: Comparing humans to automation. *Journal of experimental psychology. Applied*, 6: 104–23, 07 2000. doi: 10.1037/1076-898X.6.2.104.

[35] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. Communicating robot arm motion intent through mixed reality head-mounted displays. In Nancy M. Amato, Greg Hager, Shawna Thomas, and Miguel Torres-Torriti, editors, *Robotics Research*, pages 301–316, Cham, 2020. Springer International Publishing.

[36] S M Mizanoor Rahman and Ryojun Ikeura. Calibrating intuitive and natural human–robot interaction and performance for power-assisted heavy object manipulation using cognition-based intelligent admittance control schemes. *International Journal of Advanced Robotic Systems*, 15(4):1729881418773190, 2018. doi: 10.1177/1729881418773190.

[37] Yuhao Chen, Chizhao Yang, Yu Gu, and Boyi Hu. Influence of mobile robots on human safety perception and system productivity in wholesale and retail trade environments: A pilot study. *IEEE Transactions on Human-Machine Systems*, 52(4): 624–635, 2022. doi: 10.1109/THMS.2021.3134553.

[38] Natalia D. Kieruj and Guy Moors. Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research*, 22(3):320–342, 07 2010. ISSN 0954-2892. doi: 10.1093/ijpor/edq001.

[39] C. Harry Hui and Harry C. Triandis. Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3):296–309, 1989. doi: 10.1177/0022022189203004.

[40] Jerry W. Lee, Patricia S. Jones, Yoshimitsu Mineyama, and Xinwei Esther Zhang. Cultural differences in responses to a likert scale. *Research in Nursing & Health*, 25 (4):295–306, 2002. doi: https://doi.org/10.1002/nur.10041.

[41] Rebekka Kupffer, Susanne Frick, and Eunike Wetzel. Detecting careless responding in multidimensional forced-choice questionnaires. *Educational and Psychological Measurement*, 0(0):00131644231222420, 0. doi: 10.1177/00131644231222420.

[42] Michael Goodrich and Alan Schultz. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1:203–275, 01 2007. doi: 10.1561/1100000005.

[43] Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi. An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, 43(3): 253–270, 2008.

[44] Rachid Alami, Alin Albu-Schäffer, Antonio Bicchi, Rainer Bischoff, Raja Chatila, Alessandro De Luca, Agostino De Santis, Georges Giralt, Jérémie Guiochet, Gerd

Hirzinger, et al. Safe and dependable physical human-robot interaction in anthropic domains: State of the art and challenges. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1–16. IEEE, 2006.

[45] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. *Advanced Robotics*, 33(15-16):764–799, 2019.

[46] Bilge Mutlu, Nicholas Roy, and Selma Šabanović. Cognitive human–robot interaction. *Springer handbook of robotics*, pages 1907–1934, 2016.

[47] Sami Haddadin and Elizabeth Croft. *Physical Human–Robot Interaction*, pages 1835–1874. Springer International Publishing, Cham, 2016. ISBN 978-3-319-32552-1. doi: 10.1007/978-3-319-32552-1$_6$9.

[48] Yamato Iwamura, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 449–457, 2011. doi: 10.1145/1957656.1957816.

[49] Emrah Akin Sisbot and Rachid Alami. A human-aware manipulation planner. *IEEE Transactions on Robotics*, 28(5):1045–1057, 2012. doi: 10.1109/TRO.2012.2196303.

[50] Jane Shi, Glenn Jimmerson, Tom Pearson, and Roland Menassa. Levels of human and robot collaboration for automotive manufacturing. *Performance Metrics for Intelligent Systems (PerMIS) Workshop*, (December):95–100, 2012. doi: 10.1145/2393091.2393111.

[51] Kuka robots. `https://www.kuka.com/`.

[52] Universal robots. `https://www.universal-robots.com`.

[53] Kuka robots. `https://www.rethinkrobotics.com/`.

[54] S Aldini, MG Carmichael, and D Liu. A risk reduction framework for design of physical human-robot collaboration. In *Australasian Conference on Robotics and Automation*, 2019. Accessed: 2023/10/06.

[55] Pd iso/ts 15066:2016: Robots and robotic devices. collaborative robots. Technical standard, British Standards Institute, 2016.

[56] Chan Lee and Sehoon Oh. Development, analysis, and control of series elastic actuator-driven robot leg. *Frontiers in Neurorobotics*, 13, 2019. ISSN 1662-5218. doi: 10.3389/fnbot.2019.00017.

[57] Chee-Meng Chew, Geok-Soon Hong, and Wei Zhou. Series damper actuator: a novel force/torque control actuator. In *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, volume 2, pages 533–546 Vol. 2, 2004. doi: 10.1109/ICHR.2004.1442669.

[58] B. Vanderborght, A. Albu-Schaeffer, A. Bicchi, E. Burdet, D.G. Caldwell, R. Carloni, M. Catalano, O. Eiberger, W. Friedl, G. Ganesh, M. Garabini, M. Grebenstein, G. Grioli, S. Haddadin, H. Hoppner, A. Jafari, M. Laffranchi, D. Lefeber, F. Petit, S. Stramigioli, N. Tsagarakis, M. Van Damme, R. Van Ham, L.C. Visser, and S. Wolf. Variable impedance actuators: A review. *Robotics and Autonomous Systems*, 61(12): 1601–1614, 2013. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2013.06.009.

[59] Roni-Jussi Halme, Minna Lanz, Joni Kämäräinen, Roel Pieters, Jyrki Latokartano, and Antti Hietanen. Review of vision-based safety systems for human-robot collaboration. *Procedia CIRP*, 72:111–116, 2018. ISSN 2212-8271. doi: https://doi.org/10.1016/j.procir.2018.03.043. 51st CIRP Conference on Manufacturing Systems.

[60] Niccolò Lucci, Bakir Lacevic, Andrea Maria Zanchettin, and Paolo Rocco. Combining speed and separation monitoring with power and force limiting for safe collaborative robotics applications. *IEEE Robotics and Automation Letters*, 5(4):6121–6128, 2020.

[61] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230–253, 1997.

[62] Joseph B. Lyons and Charlene K. Stokes. Human–human reliance in the context of automation. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 54:112 – 121, 2012.

[63] Mary T. Dzindolet, Linda G. Pierce, Hall P. Beck, and Lloyd A. Dawe. The perceived utility of human and automated aids in a visual detection task. *Human Factors*, 44 (1):79–94, 2002. doi: 10.1518/0018720024494856. PMID: 12118875.

[64] Eugenio Alberdi, Andrey Povykalo, Lorenzo Strigini, and Peter Ayton. Effects of incorrect computer-aided detection (cad) output on decision-making in mammography. *Academic radiology*, 11:909–18, 09 2004. doi: 10.1016/j.acra.2004.05.012.

[65] Richard P. Will. True and false dependence on technology: Evaluation with an expert system. *Computers in Human Behavior*, 7(3):171–183, 1991. ISSN 0747-5632. doi: https://doi.org/10.1016/0747-5632(91)90006-M.

[66] Robert Molloy Indramani L. Singh and Raja Parasuraman. Individual differences in monitoring failures of automation. *The Journal of General Psychology*, 120(3): 357–373, 1993. doi: 10.1080/00221309.1993.9711153.

[67] Stephan Lewandowsky, Michael Mundy, and Gerard P.A. Tan. The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, 6:104–123, 2000. ISSN 1076898X. doi: 10.1037/1076-898X.6.2.104.

[68] Bonnie Muir and Neville Moray. Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39: 429–60, 04 1996. doi: 10.1080/00140139608964474.

[69] James Llinas, Ann Bisantz, Colin Drury, Younho Seong, and Jiun-Yin Jian. Studies and analyses of aided adversarial decision making. phase 2: Research on human trust in automation. Technical report, DTIC Document, 1998.

[70] C. Kelly, M. Boardman, P. Goillau, and E. Jeannot. Guidelines for trust in future atm systems. Technical report, European Organization for the Safety of Air Navigation, 2003. A literature review.

[71] B. D. Adams, D. J. Bryant, and R. D. Webb. Trust in teams: Literature review. Technical Report CR-2001-042, Humansystems Inc., Report to Defense and Civil Institute of Environmental Medicine, 2001.

[72] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.

[73] J. David Lewis and Andrew Weigert. Trust as a social reality. *Social forces*, 63(4): 967–985, 1985.

[74] M. Bonnie Muir. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

[75] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y.C. Chen, Ewart J. De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5):517–527, 2011. ISSN 00187208. doi: 10.1177/0018720811417254.

[76] G. Charalambous, S. Fletcher, and P. Webb. The development of a scale to evaluate trust in industrial human-robot collaboration. *International Journal of Social Robotics*, 8:193–209, 2016. doi: 10.1007/s12369-015-0333-8.

[77] Antony Tran. Robot confidence modeling and role change in physical human-robot collaboration. 2019.

[78] Neville Hogan. Impedance Control: An Approach to Manipulation: Part II—Implementation. *Journal of Dynamic Systems, Measurement, and Control*, 107 (1):8–16, 03 1985. ISSN 0022-0434. doi: 10.1115/1.3140713.

[79] Mustafa Suphi Erden and Bobby Marić. Assisting manual welding with robot. *Robotics and Computer-Integrated Manufacturing*, 27(4):818–828, 2011. ISSN 0736-5845. doi: https://doi.org/10.1016/j.rcim.2011.01.003. Conference papers of Flexible Automation and Intelligent Manufacturing.

[80] R. Riener, A. Duschau-Wicke, A. König, M. Bolliger, M. Wieser, and H. Vallery. Automation in rehabilitation: How to include the human into the loop. In Olaf Dössel and Wolfgang C. Schlegel, editors, *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, pages 180–183, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-03895-2.

[81] S. Hirche and M. Buss. Human-oriented control for haptic teleoperation. *Proceedings of the IEEE*, 100(3):623–647, 2012. doi: 10.1109/JPROC.2011.2175150.

[82] Ali Utku Pehlivan, Dylan P. Losey, and Marcia K. O'Malley. Minimal assist-as-needed controller for upper limb robotic rehabilitation. *IEEE Transactions on Robotics*, 32(1):113–124, 2016. doi: 10.1109/TRO.2015.2503726.

[83] Stefano Aldini, Avinash Singh, Marc Carmichael, Yu-Kai Wang, Dikai Liu, and Chin-Teng Lin. Prediction-error negativity to assess singularity avoidance strategies in physical human-robot collaboration. 05 2021.

[84] Marc G. Carmichael and Dikai Liu. Admittance control scheme for implementing model-based assistance-as-needed on a robot. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 870–873, 2013. doi: 10.1109/EMBC.2013.6609639.

[85] Stefano Aldini, Ashlesha Akella, Avinash K. Singh, Yu Kai Wang, Marc Carmichael, DIkai Liu, and Chin Teng Lin. Effect of mechanical resistance on cognitive conflict in physical human-robot collaboration. *Proceedings - IEEE International Conference on Robotics and Automation*, 2019-May:6137–6143, 2019. ISSN 10504729. doi: 10.1109/ICRA.2019.8793748.

[86] Huseyin Atakan Varol, Frank Sup, and Michael Goldfarb. Multiclass real-time intent recognition of a powered lower limb prosthesis. *IEEE Transactions on Biomedical Engineering*, 57(3):542–551, 2010. doi: 10.1109/TBME.2009.2034734.

[87] Gabriel Noury, Margarita Tsekeni, Vanessa Morales, Ricky Burke, Marco Palomino, and Giovanni Masala. *Experiment Protocol for Human–Robot Interaction Studies*

with Seniors with Mild Cognitive Impairments, pages 243–253. 01 2021. ISBN 978-981-15-5783-5. doi: 10.1007/978-981-15-5784-2$_2$0.

[88] National Aeronautics and Space Administration. Nasa tlx: Task load index, 2020. URL `https://humansystems.arc.nasa.gov/groups/tlx/index.php`.

[89] Tatsuya Nomura and Takayuki Kanda. On proposing the concept of robot anxiety and considering measurement of it. pages 373 – 378, 01 2003. ISBN 0-7803-8136-X. doi: 10.1109/ROMAN.2003.1251874.

[90] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1:71–81, 01 2008. doi: 10.1007/s12369-008-0001-3.

[91] Rebecca Flook, Anas Shrinah, Luc Wijnen, Kerstin Eder, Chris Melhuish, and Séverin Lemaignan. On the impact of different types of errors on trust in human-robot interaction: Are laboratory-based hri experiments trustworthy? *Interaction Studies*, 20:455–486, 11 2019. doi: 10.1075/is.18067.flo.

[92] Naomi Fitter, Mayumi Mohan, Katherine Kuchenbecker, and Michelle Johnson. Exercising with baxter: Preliminary support for assistive social-physical human-robot interaction. *Journal of NeuroEngineering and Rehabilitation*, 17, 02 2020. doi: 10.1186/s12984-020-0642-5.

[93] Małgorzata Karpińska-Krakowiak. Ratings or pairwise comparisons? an experimental study on scale usability. *Economic and Environmental Studies*, 18(2 (46)): 653–664, 2018.

[94] D. Oyserman, H. M. Coon, and M. Kemmelmeier. Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128(1):3–72, 2002. doi: 10.1037/0033-2909.128.1.3.

[95] Gerhard Meisenberg and Amandy Williams. Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44(7):1539–1550, 2008. ISSN 0191-8869. doi: https://doi.org/10.1016/j.paid.2008.01.010.

[96] Noel Pearse. Deciding on the scale granularity of response categories of likert type scales: The case of a 21-point scale. *Electronic Journal of Business Research Methods*, 9:159–171, 01 2011.

[97] Benjamin Navarro, Andrea Cherubini, Aïcha Fonte, Gérard Poisson, and Philippe Fraisse. A framework for intuitive collaboration with a mobile manipulator. pages 6293–6298, 09 2017. doi: 10.1109/IROS.2017.8206532.

[98] Sivakumar Balasubramanian, Alejandro Melendez-Calderon, Agnes Roby-Brami, and Etienne Burdet. On the analysis of movement smoothness. *Journal of Neuro-Engineering and Rehabilitation*, 12:1–11, 2015. ISSN 17430003. doi: 10.1186/s12984-015-0090-9.

[99] Neville Hogan and Dagmar Sternad. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of motor behavior*, 41:529–34, 11 2009. doi: 10.3200/35-09-004-RC.

[100] S. Cremer, S. K. Das, I. B. Wijayasinghe, D. O. Popa, and F. L. Lewis. Model-free online neuroadaptive controller with intent estimation for physical human–robot interaction. *IEEE Transactions on Robotics*, 36(1):240–253, 2020. doi: 10.1109/TRO.2019.2946721.

[101] Sugeeth Gopinathan, Sonja K. Ötting, and Jochen J. Steil. A user study on personalized stiffness control and task specificity in physical human–robot interaction. *Frontiers in Robotics and AI*, 4:58, 2017. ISSN 2296-9144. doi: 10.3389/frobt.2017.00058.

[102] Behzad Sadrfaridpour, Jenny Burke, and Yue Wang. Human and robot collaborative assembly manufacturing: Trust dynamics and control. In *RSS 2014 Workshop on Human-Robot Collaboration for Industrial Manufacturing*, 2014.

[103] Christopher D. Wickens. Multiple resources and mental workload. *Human Factors*, 50(3):449–455, 2008. ISSN 00187208. doi: 10.1518/001872008X288394.

[104] Jacob D. Dodson. The relation of strength of stimulus to rapidity of habit-formation in the kitten. *Journal of Animal Behavior*, 5:330–336.

[105] J. W. M Bertrand and H. P. G. Van Ooijen. Workload based order release and productivity: a missing link. *Production Planning & Control*, 13(7):665–678, 2002. doi: 10.1080/0953728021000026276.

[106] Saeidi Hamed. *Trust-Based Control of (Semi)Autonomous Mobile Robotic Systems*. PhD thesis, Clemson University, 2016.

[107] Marc G. Carmichael, Richardo Khonasty, Stefano Aldini, and Dikai Liu. Human preferences in using damping to manage singularities during physical human-robot collaboration. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10184–10190, 2020. doi: 10.1109/ICRA40945.2020.9197093.

[108] Behzad Sadrfaridpour, Hamed Saeidi, Yue Wang, and Jenny Burke. Modeling and control of trust in human and robot collaborative manufacturing. *AAAI Spring Symposium - Technical Report*, pages 64–70, 01 2014. doi: 10.1007/978-1-4899-7668-0_7.

[109] Hamed Saeidi and Yue Wang. Incorporating trust and self-confidence analysis in the guidance and control of (semi)autonomous mobile robotic systems. *IEEE Robotics and Automation Letters*, PP:1–1, 12 2018. doi: 10.1109/LRA.2018.2886406.

[110] H. Saeidi, J. R. Wagner, and Y. Wang. A mixed-initiative haptic teleoperation strategy for mobile robotic systems based on bidirectional computational trust analysis. *IEEE Transactions on Robotics*, 33(6):1500–1507, 2017. doi: 10.1109/TRO.2017.2718549.

[111] M. G. Carmichael, R. Khonasty, S. Aldini, and D. Liu. Human preferences in using damping to manage singularities during physical human-robot collaboration. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10184–10190, 2020. doi: 10.1109/ICRA40945.2020.9197093.

[112] Qiao Wang, Dikai Liu, Marc Carmichael, Stefano Aldini, and Chin-Teng Lin. Computational model of robot trust in human co-worker for physical human-robot collaboration. *IEEE Robotics and Automation Letters*, pages 1–1, 2022. doi: 10.1109/LRA.2022.3145957.

[113] Paul Evrard and Abderrahmane Kheddar. Homotopy switching model for dyad haptic interaction in physical collaborative tasks homotopy switching model for dyad haptic interaction in physical collaborative tasks. whc'09: World haptics-3rd joint eurohaptics conference and sym-posium on haptic interfaces for virtual environment and teleoperator systems homotopy switching model for dyad haptic interaction in physical collaborative tasks. 2009. doi: 10.1109/WHC.2009.4810879ï.

[114] Stuart K Card, Thomas P Moran, and Allen Newell. *The psychology of human-computer interaction*. Crc Press, 2018.

[115] Marc G. Carmichael, Stefano Aldini, Richardo Khonasty, Antony Tran, Christian Reeks, Dikai Liu, Kenneth J. Waldron, and Gamini Dissanayake. The anbot: An intelligent robotic co-worker for industrial abrasive blasting. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8026–8033, 2019. doi: 10.1109/IROS40897.2019.8967993.

[116] Qiao Wang, Dikai Liu, Marc Carmichael, and Chin-Teng Lin. Robot trust in human co-worker based role arbitration method for physical human-robot collaboration. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[117] Allen L. Edwards. Balanced latin-square designs in psychological research. *The American Journal of Psychology*, 64(4):598–603, 1951. ISSN 00029556.

[118] Federica Ferraguti, Chiara Talignani Landi, Lorenzo Sabattini, Marcello Bonfè, Cesare Fantuzzi, and Cristian Secchi. A variable admittance control strategy for stable

physical human–robot interaction. *The International Journal of Robotics Research*, 38(6):747–765, 2019. doi: 10.1177/0278364919840415.

[119] Ulf Bockenholt. Thresholds and intransitivities in pairwise judgments: A multilevel analysis. *Journal of Educational and Behavioral Statistics*, 26:269–282, 09 2001. doi: 10.3102/10769986026003269.

[120] A. De Luca and F. Flacco. Integrated control for phri: Collision avoidance, detection, reaction and collaboration. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 288–295, 2012. doi: 10.1109/BioRob.2012.6290917.

[121] Umut Varol, Marcos J. Navarro-Santana, Juan Antonio Valera-Calero, Sergio Antón-Ramírez, Javier Álvaro Martínez, María José Díaz-Arribas, César Fernández-de-las Peñas, and Gustavo Plaza-Manzano. Convergent validity between electromyographic muscle activity, ultrasound muscle thickness and dynamometric force measurement for assessing muscle. *Sensors*, 23(4), 2023. ISSN 1424-8220. doi: 10.3390/s23042030.

[122] D Meldrum, E Cahalane, R Conroy, D Fitzgerald, and O Hardiman. Maximum voluntary isometric contraction: reference values and clinical application. *Amyotroph Lateral Scler*, 8(1):47–55, 2007. doi: 10.1080/17482960601012491.

[123] Evan A Byrne and Raja Parasuraman. Psychophysiology and adaptive automation. *Biological Psychology*, 42(3):249–268, 1996. ISSN 0301-0511. doi: https://doi.org/10.1016/0301-0511(95)05161-9. Psychophysiology of Workload.

[124] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988. doi: https://doi.org/10.1016/S0166-4115(08)62386-9.

[125] Kazunori Akizuki and Yukari Ohashi. Measurement of functional task difficulty during motor learning: What level of difficulty corresponds to the optimal challenge point? *Human Movement Science*, 43:107–117, 2015. ISSN 0167-9457. doi: https://doi.org/10.1016/j.humov.2015.07.007.

[126] Pauline Maurice, Neville Hogan, and Dagmar Sternad. Predictability, force and (anti-)resonance in complex object control. *Journal of neurophysiology*, 120, 04 2018. doi: 10.1152/jn.00918.2017.

[127] M. S. Ryoo. Human activity prediction: Early recognition of ongoing activities from streaming videos. In *2011 International Conference on Computer Vision*, pages 1036–1043, 2011. doi: 10.1109/ICCV.2011.6126349.