Towards Enabling Human-Robot Handovers: Exploring Nonverbal Cues for Fluent Human-Robot Handovers

by

Matthew Keith Xi-Jie Pan

B.A.Sc., The University of Waterloo, 2009
M.A.Sc., The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Mechanical Engineering)

The University of British Columbia
(Vancouver)

March 2018

© Matthew Keith Xi-Jie Pan, 2018
Abstract

Fundamental human to human interactions - sharing spaces, tools, handing over objects, carrying objects together - are part of the everyday experience; for most people, the task of handing over an object to another person is a natural and seemingly effortless task. However, in the context of human-robot interaction, smooth and seamless interaction is an open problem of fundamental interest for robotics designers, integrators and users alike. This thesis explores how nonverbal cues exhibited during robot giving and receiving behaviours change how users perceive the robot, and affect the handover task. Additionally, the work also investigates how robots can recognize and interpret expressions conveyed by a human giver to infer handover intent.

Over the course of several user studies examining human-human and human-robot handovers, the role of nonverbal cues such as gaze and object orientation and how they may play a part in establishing fluency and efficiency of robot-to-human handovers are investigated. These studies provide insights into how robots can be trained through observation of human-to-human handovers. Furthermore, this thesis examines the role of nonverbal cues in the less-studied human-to-robot handover interaction. In this exploration, kinematic features from motion-captured skeleton models of a giver are used to establish the intent to handover, thereby enabling a robot to appropriately react and receive the object. Additionally, changing user perceptions, geometry and dynamics of human-to-robot handovers are explored through variation of initial pose, grasp type during, and retraction speed after handover of the robot receiver.

Findings from this thesis demonstrate that nonverbal cues such as gaze and object orientation in the case of robot-to-human handovers, and kinodynamics during
human-to-robot handovers, can significantly affect multiple aspects of the interaction including user perception, fluency, legibility, efficiency, geometry and fluidity of the handover. Using a machine learning approach, recognizing handover intent from nonverbal kinematics of the giver’s pose could also be performed effectively. Thus, the work presented in this thesis indicates that nonverbal cues can serve as a powerful medium by which details of a handover can be subtly communicated to the human partner, resulting in a more natural experience in this ubiquitous, collaborative activity.
Lay Summary

Handovers of objects play a major role in successful physical collaboration between people. The same is true for collaboration between humans and robots. The objective of this thesis is to explore how robots can both recognize and display nonverbal behaviours to facilitate the receiving and giving of objects during handovers.

In this work, nonverbal cues used by a robot such as human-inspired gaze cues, object orientation, initial pose of the robot arm, how the robot grasps the object, and the robot’s movement speed following handover were found to significantly affect user perception, fluency, legibility and efficiency of handovers. Additionally, movements of people handing over objects to other people were recorded and analyzed to allow recognition of when a person is intending to handover an object to a robot.

Based on this work, future human-robot handovers can be designed to be more efficient and fluent for both robots and humans.
Preface

The work described in Chapters 3, 4, and 5 was conducted in the Collaborative Advanced Robotics and Intelligent Systems Laboratory at the University of British Columbia, Point Grey campus. These projects and associated methods were approved by the University of British Columbia’s Research Ethics Board [certificate #H10-00503]. Projects in Chapters 6 and 7 use data collected at Disney Research Los Angeles. These projects have been approved by the Disney Research Internal Review Board [IRB number DR-IRB-Pan-2016-02] and have also been approved by the University of British Columbia’s Research Ethics Board [certificate #H10-00503].

A version of Chapter 3 has been published in the Proceedings of the Association for Computing Machinery (ACM)/Institute of Electrical and Electronics Engineers (IEEE) International Conference on Human-Robot Interaction [A Moon, DM Troniak, B Gleeson, et al. (2014) Meet Me Where I’m Gazing: How Shared Attention Gaze Affects Human-Robot Handover Timing. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - Human-Robot Interaction (HRI) 2014, Bielefeld, Germany: ACM Press, pp. 334-341.]. B Gleeson, DM Troniak, A Moon and myself were the lead investigators responsible for the concept formulation, planning and directing of the project. I was responsible for the software development used in the experimental setup, as well as conducting the experiment and performing the results analysis. The manuscript was collectively written and edited by all authors.

Parts of Chapter 4 have been published in the Proceedings of the IEEE/Robotics Society of Japan (RSJ) International Conference on Intelligent Robots and Systems (IROS) [WP Chan, MKXJ Pan, EA Croft, et al. (2015) Characterization of
handover orientations used by humans for efficient robot to human handovers. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 1-6.] WP Chan and I were jointly involved in the concept development, design, execution, and analysis of this experiment.

A version of Chapter 5 has been published in the The International Journal of Robotics Research (IJRR) [MK Pan, V Skjervøy, WP Chan, et al. (2017) Automated detection of handovers using kinematic features. The International Journal of Robotics Research, SAGE Publications, UK: London, England 36(5-7): 721-738.] I was the lead investigator in this research, responsible for all major areas of concept formation, data collection and analysis, as well as manuscript composition. WP Chan and I were jointly involved in the design and execution of this experiment. Data processing was performed with the assistance of V Skjervøy.

Portions of Chapter 6 have appeared in the Human-Robot Interaction in Collaborative Manufacturing Environments Workshop at the IEEE/RSJ International Conference on Intelligent Robots and Systems [MKXJ Pan, EA Croft, and G Niemeyer. (2017) Validation of the Robot Social Attributes Scale (RoSAS) for Human-Robot Interaction through a Robot-to-Human Handover Use Case. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) - Workshops, IEEE]. Additionally, a manuscript closely resembling the contents of Chapter 6 has also been presented at the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018) held in Chicago, IL, USA from March 5-8, 2018 and appears in the conference proceedings. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis, as well as the majority of manuscript composition. This experimental study was conducted at Disney Research Los Angeles (DRLA) under the supervision of G Niemeyer. Analysis of the data was performed at University of British Columbia (UBC) under the supervision of both EA Croft and G Niemeyer (remotely).

A version of Chapter 7 has been included in the proceedings of the IEEE Haptics Symposium (HAPTICS) 2018 held in San Francisco, California, USA from March 25-28, 2018. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis, as well as the majority of the manuscript’s composition. This chapter’s contents were derived from the same
study found in Chapter 6; the experimental study was conducted at DRLA under the supervision of G Niemeyer. Analysis of the data was performed at UBC under the supervision of both EA Croft and G Niemeyer (remotely).

The composition of this thesis conforms to the guidelines for structure and format of UBC theses and dissertations laid out by the UBC Faculty of Graduate and Postdoctoral Studies which can be found in [117].
Table of Contents

Abstract ............................................................... ii
Lay Summary ......................................................... iv
Preface ............................................................... v
Table of Contents ...................................................... viii
List of Tables ......................................................... xiv
List of Figures ......................................................... xvi
Glossary .............................................................. xix
Acknowledgements ................................................... xxii

1 Introduction ......................................................... 1
  1.1 Exploring Handovers ............................................. 5
  1.2 Research Objective .............................................. 6
  1.3 Thesis Organization ............................................. 8
    1.3.1 Robot-to-Human Handovers ............................... 9  
    1.3.2 Human-to-Robot Handovers ............................. 10
  1.4 Thesis Notes ................................................... 12
  1.5 Contributions ................................................ 13

2 Background ......................................................... 14
3 The Effect of Robot Gaze in Robot-to-Human Handovers

3.1 Introduction

3.2 Background - Gaze in Handovers

3.3 Study I: Observing Gaze Patterns in Human-to-Human Handovers
   3.3.1 Experimental Procedure
   3.3.2 Results
   3.3.3 Discussion

3.4 Study II: Impact of Human-Inspired Gaze Cues on First-Time Robot-to-Human Handovers
   3.4.1 Physical Handover Cues
   3.4.2 Experimental Gaze Cues
   3.4.3 Experimental Procedure

3.5 Technical Implementation

3.6 Results
   3.6.1 Handover Timing
   3.6.2 Subjective Experience

3.7 Discussion

3.8 Conclusions
   3.8.1 Limitations
   3.8.2 Future Work

4 Characterization of Handover Object Orientations Informing Efficient Robot-to-Human Handovers

4.1 Introduction

4.2 Background

4.3 Objectives

4.4 Object Handover User Study
4.4.1 Experiment Design ........................................ 52
4.4.2 Objects ..................................................... 53
4.4.3 Participants ............................................... 55
4.4.4 Participant Task ........................................... 55
4.4.5 Motion-Capture System .................................. 57
4.4.6 Hypotheses .................................................. 59

4.5 Data Analyses ................................................. 59
4.5.1 Handover Orientation Extraction .......................... 59
4.5.2 Affordance Axes .......................................... 62
4.5.3 Patterns in Handover Orientations ....................... 62
4.5.4 Comparison of Handover Orientations Across Conditions . 63

4.6 Results .......................................................... 64
4.6.1 Handover Orientation and Affordance Axis .............. 64
4.6.2 Patterns in Handover Orientations ....................... 67
4.6.3 Comparison of Handover Orientations Across Conditions . 69

4.7 Discussion ....................................................... 71
4.7.1 Patterns in Handover Orientations ....................... 71
4.7.2 Comparison of Handover Orientations Across Conditions . 71
4.7.3 Implications Towards Building Intelligent Robots ...... 73

4.8 Conclusions .................................................... 74

5 Automated Detection of Handovers using Kinematic Features ... 76
5.1 Introduction ....................................................... 78
5.2 Prior Work ....................................................... 79
5.2.1 Non-Verbal Cues, Proxemics and Kinematics in Handovers 79
5.2.2 Machine Learning for Handovers ........................ 80
5.3 Objectives and Approach ..................................... 82
5.4 Experimental Setup .......................................... 83
5.5 Data Analysis ................................................... 83
5.5.1 Data Post-Processing ..................................... 83
5.5.2 Feature Extraction ........................................ 84
5.5.3 Labelling ..................................................... 86
5.5.4 Test Set Generation ....................................... 87
6.6 Discussion ........................................ 134
6.6.1 ROSAS Internal Consistency and Dimensionality .... 134
6.6.2 Effect of Conditions ................................ 135
6.6.3 Effect of Repeated Interaction Over Time ............. 137
6.6.4 Implications for HRI ............................... 137
6.7 Conclusions ...................................... 138
6.7.1 Limitations ................................. 139
6.7.2 Future Work ................................. 139

7 Exploration of Geometry and Forces Occurring Within Human-to-
Robot Handovers ................................... 141
7.1 Introduction .................................... 142
7.2 Experimental Design ............................. 143
7.3 Geometry .................................. 143
  7.3.1 Results .................................... 144
  7.3.2 Discussion ................................. 145
7.4 Dynamics .................................... 147
  7.4.1 Results .................................... 148
  7.4.2 Discussion ................................. 149
7.5 Learning .................................... 151
  7.5.1 Results .................................... 153
  7.5.2 Discussion ................................. 153
7.6 Conclusions .................................... 154
  7.6.1 Future Work ................................. 155

8 Conclusions ........................................ 156
8.1 Conveying Non-Verbal Cues During Handover .............. 157
  8.1.1 Summary of Findings .......................... 157
  8.1.2 Implications ............................... 161
  8.1.3 Limitations ............................... 161
  8.1.4 Future Work ............................... 162
8.2 Recognizing Non-Verbal Cues from Humans ............... 163
  8.2.1 Summary of Findings ........................ 163
8.2.2 Implications .............................................. 165
8.2.3 Limitations .............................................. 165
8.2.4 Future work ............................................ 166

Bibliography .................................................... 167

A Supporting Materials for Investigating Robot Gaze in Robot-to-Human Handovers .............................................. 184
A.1 Study Advertisements ..................................... 184
A.1.1 Poster ..................................................... 185
A.1.2 E-Mail ..................................................... 186
A.2 Consent ...................................................... 187
A.2.1 Written Consent ......................................... 187
A.2.2 Verbal Consent ......................................... 189
A.3 Surveys ....................................................... 190
A.3.1 Single Condition Survey ................................ 190
A.3.2 Condition Comparison Survey ......................... 191

B Supporting Materials for Investigation of Handover Object Orientations and Automated Detection of Handovers ................................. 192
B.1 Study Advertisements ..................................... 192
B.1.1 Poster ..................................................... 193
B.1.2 E-Mail ..................................................... 194
B.2 Consent ...................................................... 195

C Supporting Materials for Investigating Human-to-Robot Handovers .............................................. 197
C.1 Consent ....................................................... 197
C.2 Survey ....................................................... 203
# List of Tables

| Table 1.1 | Non-exhaustive table of generalized comparisons of strengths and weaknesses between robots and humans. | 4 |
| Table 3.1 | Table of questionnaire results in the robot-to-human handover gaze study. | 41 |
| Table 4.1 | Results of comparison of handover orientations between conditions in the handover object orientation study. | 70 |
| Table 5.1 | Features extracted from kinematic data obtained from the motion capture system during the study investigating the automated detection of handovers. | 86 |
| Table 5.2 | Training and testing data specifications for the automated detection of handovers. | 92 |
| Table 5.3 | The 22 features selected for use as the Support Vector Machine (SVM) input space to automatically detect handovers and their predictive ability rankings. | 94 |
| Table 5.4 | Optimized hyperparameter values and corresponding cross-validation error for the SVMs using various kernels. | 100 |
| Table 5.5 | Confusion matrix for handover detection using a SVM with various kernels on motion capture data. | 102 |
| Table 5.6 | Pilot study confusion matrix for handover detection using a SVM with Radial Basis Function (RBF) kernel applied to features generated via a Kinect version 2 sensor. | 110 |
Table 6.1  Table of ROSAS items categorized by the factors competence, warmth and discomfort. ........................................ 121
Table 6.2  Table of experimental conditions for the human-to-robot handover study. .................................................. 128
Table 7.1  Comparison of Cartesian positions of handover objects between human-robot and human-human handovers. ..... 146
Table 8.1  Summary table of robotic nonverbal cues studied in this thesis and their observed impacts. .............................. 158
List of Figures

Figure 3.1  Timeline of giver’s gaze patterns observed from human-human handovers. ........................................... 27
Figure 3.2  Demonstration of two frequently observed gaze behaviours from the human-human handover study. ............... 29
Figure 3.3  Illustration demonstrating experimental set-up and gaze for each condition in the robot-to-human handover gaze study. ................. 32
Figure 3.4  Diagram depicting the gaze cues for the No Gaze, Shared Attention, and Turn-Taking gazes. .............................. 34
Figure 3.5  System flow diagram describing algorithm used for PR2 robot gaze during handovers. .................................................. 38
Figure 3.6  Chart of handover timing results for various gaze conditions in the robot-to-human handover study. .................... 39

Figure 4.1  Illustration depicting the importance of considering object orientation during robot-to-human handovers. ............... 49
Figure 4.2  Common everyday objects affixed with retro-reflective markers that were used in the handover object orientation user study. .......... 54
Figure 4.3  Experimental setup for examining object orientations during handovers. ................................................................. 56
Figure 4.4  Diagram of retroreflective marker placement and labels on participants during the study investigating the automated detection of handovers. ......................................................... 58
Figure 4.5  Charts representing typical hand trajectories of giver and receiver. ................................................................. 60
Figure 4.6  Base coordinate frame used during examination of object orientations during handovers as defined by giver and receiver locations. .......................................................... 61

Figure 4.7  Monte Carlo simulation results showing histograms of the angle between the mean affordance axis for each object and the affordance axis of the object rotated by the orientation of that object used by each participant. ........................................ 63

Figure 4.8  Figure showing orientation of a teapot for one and multiple trials as it is being handed over from giver to receiver. .... 65

Figure 4.9  Handover orientations employed during human-human handover for all objects in all conditions. ...................... 66

Figure 4.10 Histograms of the angles between $\hat{R}$ $\phi_{\text{Aff}}$ and $R_{i}$ $\hat{\phi}_{\text{Aff}}$ (i.e., $\theta_{i}$) for each object in condition C during the study examining handover orientations. .......................... 68

Figure 5.1  Diagram showing a giver passing a book to the receiver, and the corresponding motion-captured scene. ................. 85

Figure 5.2  Charts depicting resubstitution loss, generalization error and out-of-bag error for Gini and entropy/information gain split criterion as a function of bootstrap sampling fraction. ....... 89

Figure 5.3  Charts depicting resubstitution loss, generalization error and out-of-bag error for Gini and entropy/information gain split criterion as a function of number of learners. .................... 89

Figure 5.4  Plot of the SVM feature space arranged by importance scores generated from the bagged random forest. .............. 96

Figure 5.5  Parallel coordinates plot of the input space of the handover classifier determined by using a bagged random forest. ...... 97

Figure 5.6  Kernel search spaces for optimization of the regularization parameter (C) for the handover classifying SVM .......... 99

Figure 5.7  Diagram of joints and labels as automatically generated by the Microsoft Kinect version 2 sensor. ....................... 108

Figure 6.1  Experimental setup for human-to-robot handover experiment. 123
Figure 6.2 Flowchart of *quick* (top row) and *mating* grasp types used during the human-to-robot handover experiment.  

Figure 6.3 Scree plots for ROSAS factors to examine dimensionality of items.  

Figure 6.4 Participants ratings of the robot’s competence, warmth and discomfort over conditions as reported during the human-to-robot handover study.  

Figure 6.5 Interaction plot showing participants ratings of the robot’s discomfort based on grasp type at levels of retraction speed.  

Figure 6.6 Participants ratings of the robot’s competence, warmth and discomfort over time as recorded during the human-to-robot handover study.  

Figure 7.1 Scatterplots of initial object positions (from the overhead perspective) for both the *down* and *up* initial arm pose conditions presented during the human-to-robot handover study.  

Figure 7.2 Scatterplots of initial object azimuth and elevation angles for both the *down* and *up* initial arm pose conditions presented during the human-to-robot handover study.  

Figure 7.3 Sample end effector axial force vs. time plot for a participant depicting features used in the dynamics analysis of the human-to-robot handover study. Negative forces indicate pushing (compressive) forces exerted against the end effector whereas positive forces indicate pulling (tension) forces.  

Figure 7.4 Maximum absolute and relative retraction forces as experienced by the giver in the human-to-robot handover study.  

Figure 7.5 Maximal forces during handover as experienced or applied by givers during the human-to-robot handover study over trials ordered chronologically.
# Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance, a set of statistical techniques to identify sources of variability between groups</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>DMP</td>
<td>Dynamic Movement Primitive</td>
</tr>
<tr>
<td>GSS</td>
<td>Golden Section Search</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>HHI</td>
<td>Human-Human Interaction</td>
</tr>
<tr>
<td>HRI</td>
<td>Human-Robot Interaction</td>
</tr>
<tr>
<td>IIWA</td>
<td>Intelligent Industrial Work Assistant used to describe a series of robot arms developed by KUKA Robotics designed for human-robot collaborative tasks</td>
</tr>
<tr>
<td>LBR</td>
<td>Leichtbauroboter, German for lightweight robot - used to describe a series of robot arms developed by KUKA Robotics designed for human-robot collaborative tasks</td>
</tr>
<tr>
<td>MANOVA</td>
<td>Multivariate Analysis of Variance, an ANOVA with several dependent variables</td>
</tr>
<tr>
<td>PR2</td>
<td>Personal Robot 2, a robot platform developed by Willow Garage</td>
</tr>
<tr>
<td>ROS</td>
<td>Robot Operating System</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial Basis Function</td>
</tr>
<tr>
<td>ROSAS</td>
<td>Robotic Social Attributes Scale</td>
</tr>
<tr>
<td>SPI</td>
<td>Successive Parabolic Interpolation</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
</tbody>
</table>
Acknowledgements

I offer my enduring gratitude to several people and organizations who have made significant contributions not only to this thesis, but to my knowledge of the scientific world:

- My supervisor Prof. Elizabeth A. Croft, who has graciously accepted me as graduate student and have provided me with support and guidance throughout my time as a masters and doctoral student,

- Dr. Günter D. Niemeyer and Lanny Smoot who have encouraged and supported me in my forays into mixing virtual reality and human-robot interaction during my internship at DRLA,

- My colleagues and friends at the Collaborative Advanced Robotics and Intelligent Systems (CARIS) Laboratory for always being there to exchange ideas and over lunches, dinners and spontaneous meetings,

- The National Sciences and Engineering Research Council of Canada (NSERC) for providing me with funding that was critical in carrying out this research,

- My parents Danny and Tina Pan, the people who never stopped believing in my abilities as a scientist; and lastly,

- My wife and best friend, Bernice Li, for encouraging me to reach farther and dream bigger throughout my graduate career.

I would like to express additional thanks to Karon MacLean, Wesley Chan, Vinay Chawda, Winston Yao, Vidar Skervøy, AJung Moon, Brian Gleeson, Minhua
Zheng, Benjamin Blumer, Jonathan Benitez and Alisa Wyman for their support on various parts of the work presented in this thesis.
Dedicated to my loving wife Bernice and daughter Maven.
Chapter 1

Introduction

The worldwide demand for robotics technologies has been steadily growing over the last two decades to meet the needs of industrial and service industries [1, 57]. There is little doubt that this trend will persist, at the very least, as robotics technologies become sufficiently advanced to enter into new areas of service. The International Federation of Robotics (IFR) predicts that the demand for domestic and service robots, for example, is expected to increase six-fold within the next two years alone [57] as the abilities of robots mature to address growing opportunities and challenges in various markets and sectors. They predict that future technology development will allow robots to:

- deploy in growing consumer markets and decreased product life cycles which require expansion of output capabilities for competitiveness
- provide a solution for a significant workforce gap created by an aging and retiring population
- address the health care needs of this aging population
- perform work considered tedious, hazardous and/or unsanitary to human workers

With advancing technological capabilities and increasing evidence that robotics will be able to serve in such capacities, the use of robots in the field has expanded
greatly - from initially being used in heavy industries (such as automotive manufacturing) in the 1960s, they are now seen performing tasks such as warehouse logistics [e.g., 7], home vacuum cleaning [e.g., 59], customer service [e.g., 108] and laparoscopic surgery [e.g., 58]. The increasing pervasiveness of robotics in these environments capitalizes on rapidly developing technologies to deliver convenience, cost-savings, and/or higher efficiencies. Many organizations adopting robotic technologies see this as a way to lay the foundations for innovative and sustainable development, and to improve consistency and efficiency of how products are produced and services rendered. In many of these applications, the predominant model of how robots are deployed is to have them completely replace human workers. For many dangerous, repetitive, precise, and simple tasks which may adversely affect human workers (either mentally or physically), the justification for using this model is sound: robots are undeniably useful for rote operations which may involve handling heavy/dangerous materials or require precise positioning. Such tasks involve operations that humans are neither good at or designed for.

However, there are abilities that robots continue to lack or under-perform in when compared to humans. A recent article published by the Institute of Electrical and Electronics Engineers (IEEE) Spectrum Magazine entitled “Shockingly, Robots Are Really Bad at Waiting Tables” highlights some of the problems when current robot technologies attempt to replace humans though do not have the necessary skill-sets [2]. The article describes two restaurants in China which gained notoriety for using robotic waiters to deliver food to people, a scheme which was scrapped shortly after the robots’ introduction. One restaurant employee was quoted as saying that “their skills are somewhat limited...they can’t take orders or pour hot water for customers.” Indeed, the robots used in the restaurants were only able to travel on pre-planned paths and carry trays of food to the vicinity of tables. Ackerman writes: “those are just two of the many, many more skills that human servers have, because it’s necessary to have many, many more skills than this to be a good server”. Compared to humans, robots currently under-perform in other areas including object recognition, flexibility in decision-making, understanding natural-language queries, and dexterity. If a model of robots being used to replace humans is blindly adopted without considering the strengths and weaknesses of the robot, situations such as those experienced with the robot waiters will occur.
Given that robotics is still a developing technology, many sectors are investing in another model of robotics which mitigate the aforementioned problems: instead of having robots replace humans, have robots work alongside humans to have a complement of capabilities which negate the weaknesses of both. Table 1.1 shows some comparisons of select strengths and weaknesses of both, and aims to highlight how collaboration within human-robot teams may significantly increase task proficiency over the robot or human individually. In this collaboration model, humans and robots actively collaborate in close proximity to each other to complete tasks. For example, in a manufacturing scenario, a robot could assist a human worker in holding up a heavy part while the worker secures the part to an assembly using fasteners. Also, a robot could hand over tools to a worker at the right time to enable the worker to perform certain tasks, much like how a nurse hands over surgical instruments to a surgeon during an operation. In these situations, the purpose of the robot is to enable, support, and enhance the capabilities of the human worker through Human-Robot Interaction (HRI). Alternatively, the reverse will be true equally often - for example, the human may be instructed by the robot to provide dexterity in a job that the robot can otherwise handle. In this case where the robot is managing the overall task, it can be argued that the human is supporting and augmenting the capabilities of the robot. In both these scenarios, however, it is the co-operation and collaboration of both human and robotic agents which leads to a more versatile work unit.

To aid in deploying human-robot collaborative systems, robotic manufacturers are beginning to provide platforms designed to be used for HRI; for example, KUKA Robotics Corporation (Augsburg, Germany) have recently released their LBR IIWA lightweight robot which they describe as the “world’s first series-produced sensitive, and therefore human-robot collaborative compatible, robot” [70]. Other robotics-centered organizations such as Rethink (Boston, MA), ABB (Zürich, Switzerland), Franka Emika (Munich, Germany) are also offering robots that specialize in HRI. Users of robotics technologies are also investing in the human-robot collaborative model: General Motors has partnered with several academic institutions in a research project called Collaborative, Human-focused, Assistive Robotics for Manufacturing (CHARM) which explored ways of deploying
### Table 1.1: Non-exhaustive table of generalized comparisons of strengths and weaknesses between robots and humans.

<table>
<thead>
<tr>
<th>Robots</th>
<th>Humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow impedance range</td>
<td>Wide impedance range</td>
</tr>
<tr>
<td>Large payload capacity</td>
<td>Small payload capacity</td>
</tr>
<tr>
<td>Precise, repeatable actuation</td>
<td>Non-precise, semi-repeatable actuation (can be affected by factors such as disease, muscle weakness, etc.)</td>
</tr>
<tr>
<td>Limited flexibility (added flexibility is costly)</td>
<td>Very flexible</td>
</tr>
<tr>
<td>Limited dexterity (added dexterity is costly)</td>
<td>Very dexterous</td>
</tr>
<tr>
<td>Able to accurately plan and keep track of place in multi-step procedures</td>
<td>Prone to mistakes in executing ordered procedures and/or inefficient planning of tasks</td>
</tr>
<tr>
<td>Sensor specialization and access to database information allows for accurate and timely quality assessment</td>
<td>Quality assessment methods generally limited by trade-offs (e.g., crude and fast or comprehensive and slow)</td>
</tr>
<tr>
<td>Limited ability to self-diagnose issues affecting task performance</td>
<td>Able to investigate and establish causality to rectify issues affecting task performance</td>
</tr>
</tbody>
</table>

robots in manufacturing environments to collaborate with and enhance workers’ capabilities while making production lines more flexible [40].

As service robots are deployed in environments that permit close proximity to human users for human-robot collaboration, it is essential that critical capabilities in HRI are developed. In particular, in order for robots to interact effectively and naturally with human counterparts, their behaviours must be carefully designed and informed by the recognition of cues provided by the human and environmental elements. Likewise, dexterous manipulations performed by the robot should be perceived as safe, reliable and predictable in context of the situation and/or environment. In other words, cooperative interaction between robots and people must
consider how the robots perceive the people and environments they are interacting with, and how the robots themselves are perceived.

This thesis explores these concepts by examining one type of fundamentally necessary interaction between robots and humans during physical cooperation: object handovers. The ability of robots to safely and effectively handover objects is a crucial capability for collaborative human-robot interaction. It is a necessary skill for robots if they are to be able to fully assist and cooperate with human users within shared task environments. Revisiting the example of waiter robots deployed in restaurants, if the robot were able to appropriately handover dishes and glasses, the robot would arguably be a much more effective waiter (granted, the robot would also need additional skills to be able to completely replace human waiters). Giving or receiving an object is a skill that will allow robots to be increasingly useful in contexts such as manufacturing and assistive care. For example, a robot can handover or retrieve a tool to and from a line worker in a manufacturing plant depending on which tool the worker needs at the time, increasing worker productivity. In an assisted living situation, a robotic assistant can help fetch a TV remote or drink for a mobility-limited user. In these and other use cases, the ability for the robot to effectively participate in a handover adds immense value for interactive robots that will increase their ubiquity within the home and workplace.

1.1 Exploring Handovers

The act of handing over objects is perhaps the quintessential representation of cooperation between people. It is a necessary ability that is conducted habitually and naturally everyday; it is overlooked by most as a rudimentary skill that occurs almost reflexively. However, when carefully examining the coordination involved in handovers, it is apparent that they are intricate interactions involving multiple subtle, nonverbal cues generated and interpreted by both agents. Object handovers require kinodynamic (kinematic + dynamic) negotiation, gesture and motion prediction, coordinated hand-eye motions, and tactile sensing on the part of both participants - all within a matter of seconds.

Examining and considering these nonverbal cues is especially important for a developing a robot assistant that can give and receive objects to a person: to
have a robot smoothly and seamlessly give or receive an object is a technical and, arguably, an artistic challenge of being able to generate, interpret and react to nonverbal cues. This presents an open problem of designing, implementing and integrating several sub-systems which allow a robot to sense and appropriately react before, during and after object handovers. Specifically, these sub-systems should provide the robot with the ability to recognize and understand the underlying kinematic and dynamic interactions between the giver and receiver, with the intention of and participating in this negotiation in a fluent and efficient manner. In the example of a robot giver passing an object to a human, this includes having the robot provide clear and understandable cues to indicate that it wants to hand over an object, fluently negotiate where and when the handover is to occur, recognize when the human user has a stable grasp of the object, and then appropriately release its own grasp on the object before safely retracting its manipulator. For the inverse situation where a robot receives an object from a human, the robot must interpret the human’s nonverbal cues that he/she wants to hand over an object, fluently negotiate where and when the handover is to occur, obtain and maintain a stable grasp the object, and provide cues to effectively indicate to the human that it achieved a stable grasp. Both directions of handover (human-to-robot and robot-to-human) are examined in this thesis.

Considering these challenges, it is apparent that handovers present an exemplar and useful context in which to investigate robot behaviours that support HRI. They are an interaction which can be used as a model to study other HRIS. From this perspective, the broad aim of this research is to contribute to the understanding of HRI design with respect to implementation of simple, everyday interaction tasks.

1.2 Research Objective

This thesis explores the roles that a handful of nonverbal cues play in human-robot handovers and the effects of manipulating these cues on a robot that acts as both giver and receiver of objects. This work does not attempt to map the boundaries of the design space that is available to robot designers, integrators and users for developing handover gestures, nor does it limit exploration to human-like methods of conducting handovers. Rather, it explores how some nonverbal cues exhibit-
ited by a human can be recognized and leveraged during the handover negotiation and how robots can use nonverbal cues to affect the interaction and how peoples’ perceptions of them. The goal of this body of work is to provide some indication on how fluent and efficient human-robot handovers can be achieved through the use of nonverbal cuing and interaction. In particular, the nonverbal cues that are examined include: robot gaze, object alignment/orientation, proxemics, robot grasping, kinematics, and dynamics. The primary aim of this thesis is to convey that just like human-human handovers, the negotiation that occurs in achieving fluent and efficient human-robot handovers can be deceptively complex and relies on many nonverbal cues.

The corpus of work contained in this thesis primarily centers around two research questions:

Q1 How do nonverbal cues exhibited by the robot during robot giving and receiving behaviours change how users perceive the robot, and affect the handover negotiation?

Q2 How can a robot adequately recognize and interpret nonverbal cues conveyed by a human to infer object attributes as well as handover intent?

To address the first question, the effects that gaze cues generated by a robot giver (with an articulated head) have on handover interactions are considered (Chapter 3). In the case where a robot arm acts as a receiver in the handover negotiation, this thesis examines the outcomes of manipulating seemingly innocuous factors such as initial arm pose prior to, grasp method during, and object retraction following object handover changes people social perceptions of the robot (Chapter 6), as well as the kinodynamics of the interaction (Chapter 7).

To examine the premise of the second research question, a portion of this work has been dedicated to explore how robots can be trained to pass objects to human receivers while considering the object’s affordances, such that an object can be passed to a human receiver in a manner which is most comfortable to the receiver (Chapter 4). In this study, the viability of using human-human handover data sets for this task is examined. In a separate study, how user poses and kinematics can be leveraged through motion tracking systems to detect a human giver’s intent to handover through machine learning are considered (Chapter 5).
A more detailed summary of each of these works appears in the following section.

1.3 Thesis Organization

The investigations on human-robot handovers contained within this thesis can be considered to span two different sub-domains or tasks: robot-to-human handovers and human-to-robot handovers. At first glance, it may appear that there is very little distinction between these two categories since both have robot and human agents participating in a single type of activity. However, from the perspective of the robot, these tasks are very different: robot-to-human handovers require the robot to act as an object giver, whereas human-to-robot handovers employ the robot as an object receiver - these activities encompass very different technical challenges.

For robot-to-human handovers, where the robot acts as an object giver (assuming that the giver initiates and receiver is ready to participate in the handover), the robot is required to:

1. grasp the handover object safely and securely prior to handover;
2. effectively communicate to a human counterpart that it wants to handover an object;
3. consider how convenient it is for the human receiver to physically grasp the object, and orient objects such that their affordances are directed towards the receiver;
4. dynamically negotiate with the human receiver to perform the handover in an expeditious and safe (in terms of the object) manner; and
5. recognize when the receiver has a stable grasp of the object before tactfully releasing its own grasp to complete the exchange.

On the other hand, in human-to-robot handovers, the robot receiver must:

1. recognize that a human is initiating a handover;
2. detect when and where the handover will take place;

3. consider how to grasp the object; and

4. communicate to the human giver that it ‘has’ the object.

Due to these very distinct sets of abilities required by the robot givers and receivers, this thesis has been organized into two parts: the first part (Chapters 3 and 4) contains work on robot-to-human handover, whereas the second part (Chapters 5 to 7) examines human-to-robot handovers.

1.3.1 Robot-to-Human Handovers

Chapter 3 provides empirical evidence that using human-like gaze cues during human-robot handovers can improve the timing and perceived quality of the handover event. Fluent, legible handover interactions require appropriate nonverbal cues to signal handover intent, location and timing. Based upon observations of human-human handovers, gaze behaviours were implemented on a PR2 humanoid robot. The robot handed over water bottles to a total of 102 inexperienced subjects while varying its gaze behaviour: no gaze, gaze designed to elicit shared attention at the handover location, and the shared attention gaze complemented with a turn-taking cue. Subject perception of and reaction time to the robot-initiated handovers were compared across the three gaze conditions. Results indicate that subjects reach for the offered object significantly earlier when a robot provides a shared attention gaze cue during a handover. A statistical trend of subjects preferring handovers with turn-taking gaze cues over the other conditions was also observed. The work presented in Chapter 3 demonstrates that gaze can play a key role in improving user experience of human-robot handovers, and help make handovers fast and fluent. Within the context of this thesis, the inclusion of this manuscript assists in highlighting that subtle nonverbal cues profoundly impact the
way that handover tasks are carried out and can affect user experience of the task.

In Chapter 4, steps are taken towards enabling robot givers to naturally pass objects to human receivers comfortably by considering the object’s affordances. In line with the first research question investigated in this thesis and towards enabling robots to learn proper handover object orientations during handover, this work seeks to determine if the nonverbal cue of natural object orientation observed during human-to-human handovers can be used to train robots to consider handover object affordances. By observing human-human handovers, natural handover object orientations are compared with giver-centered and receiver-centered handover alignments for twenty common objects. A distance minimization approach was used to compute mean handover orientations. It is posited that computed means of receiver-centered orientations could be used by robot givers to achieve more efficient and socially acceptable handovers. Furthermore, the notion of affordance axes for comparing handover alignments is introduced in this work, and offers a definition for computing them. As a result of this study, observable patterns were found in receiver-centered handover orientations. Comparisons show that depending on the object, natural handover orientations may not be receiver-centered; thus, robots may need to distinguish between good and bad handover orientations when learning from natural handovers.

1.3.2 Human-to-Robot Handovers

As opposed to the wide body of work conducted in robot-to-human handovers, human-to-robot handovers is a domain that, at the time of writing, remains largely unexplored. The work presented in Chapter 3 was the result of a collaboration between several organizations and individuals, including myself. As mentioned in the preface to this thesis, B Gleeson, DM Troniak, A Moon and myself were the lead investigators responsible for the concept formulation, planning and directing of the project. I was responsible for the software development used in the experimental setup, as well as conducting the experiment and performing the results analysis. The manuscript was collectively written and edited by all authors. In addition to this work’s inclusion in this thesis on nonverbal cues during handovers, it also appears in A Moon’s thesis where she focuses on the interweaving of cues and plans between humans and robots [87].

The work presented in Chapter 4 was the result of a collaboration between WP Chan and myself. WP Chan and I were jointly involved in the concept development, design, execution, and analysis of this experiment.
explored. Thus, the work presented in this thesis within this domain focuses on developing building blocks for robot receiving.

As one of these building blocks, Chapter 5 details work conducted to have a robot recognize that a human giver is initiating an object handover. In this project, the use of kinematic motions recognized by Support Vector Machines (SVMs) for the automatic detection of object handovers from the perspective of an object receiver. The classifier uses the giver’s kinematic behaviours (e.g., joint angles, distances of joints from each other and with respect to the receiver) to determine a giver’s intent to handover an object. A bagged random forest was used to determine how informative features were in predicting the occurrence of handovers, and to assist in selecting a core set of features to be used by the classifier. Altogether, 22 kinematic features were chosen for developing handover detection classification models. Test results indicated an overall maximum accuracy of 97.5% by the SVM in its capacity to distinguish between handover and non-handover motions. The classification ability of the SVM was found to be unaffected across four kernel functions (linear, quadratic, cubic and radial basis). These results directly address the 2nd research posed by this thesis and demonstrate considerable potential for detection of handovers and other gestures for human-robot interaction using kinematic features.3

Following this work, a robot receiving system was developed using a KUKA LBR iiWA 7 R800 robot arm (KUKA, Augsburg, Germany). Chapters 6 and 7, leverage this system to examine how users behave and perceive robots in human-to-robot handovers. In Chapter 6, users’ social perceptions of robots within the domain of human-to-robot handovers are explored. Using the Robotic Social Attributes Scale (ROSAS), developed by Carpinella et al. in [25], the work explores how users socially judge robot receivers as three kinodynamic factors are varied: initial position of the robot arm prior to handover (up vs. down), grasp method employed by the robot when receiving a handover object trading off perceived object safety for time efficiency or vice versa, and retraction speed of the arm following

---

3I was the lead investigator in the research presented in Chapter 5, responsible for all major areas of concept formation, data collection and analysis, as well as manuscript composition. WP Chan and I were jointly involved in the design and execution of this experiment. Data processing was performed with the assistance of V Skjervøy.
handover (slow vs. fast). The results show that over multiple handover interactions with the robot, users gradually perceive the robot receiver as being less discom-forting and having more emotional warmth. Additionally, we have found that by varying grasp method and retraction speed, users may hold significantly different judgments of robot competence and discomfort. These results provide empirical evidence that users are able to develop social perceptions of robots which can change through modification of robot nonverbal receiving behaviours and through repeated interaction with the robot, addressing the first question posed in this thesis.

In Chapter 7, kinematic and dynamic data recorded from human-to-robot handover trials are examined. Analysis of the kinodonamic data shows that the robot’s initial pose can inform the giver about the upcoming handover geometry and impact fluency and efficiency. Also, variations in grasp method and retraction speed appear induce significantly different interaction forces. This effect may occur by changing the giver’s perception of object safety and hence their release timing. Alternatively, it may stem from unnatural or mismatched robot movements. Additionally, the findings indicate that making the robot predictable is important: a learning effect within the forces applied to the handover object linearly decline over repeated trials. Simultaneously, as shown in Chapter 6, the participants’ self-reported discomfort with the robot decreases and perception of emotional warmth increases. Thus, it is posited that users are learning to predict the robot, becoming more familiar with its behaviours, and perhaps becoming more trusting of the robot’s ability to safely receive the object. As a result, these findings suggest that a robot can become a trusted partner in collaborative tasks.

1.4 Thesis Notes

In this thesis, consistent terminology will be used when referring to participants involved in an object handover, regardless if whether the participant is a human or robot. The participant that initially holds the object and transfers the object to the other participant through the handover interaction will be referred to as the giver. The participant that the giver transfers the object to during the handover interaction will be referred to as the receiver.
Additionally, to narrow the scope of this thesis, the work contained within this thesis adopts a model of object handover that is most commonly found and used in literature (presented in Chapter 2). In this model, the following is assumed or enforced:

- Handovers are conducted nonverbally between participants.
- The giver does not consider readiness of the receiver to participate in the handover.
- The giver always initiates the handover interaction and waits for the receiver to take the object.

1.5 Contributions

To achieve the objectives of the research described in this thesis, the following contributions were made:

- Experimental results of how nonverbal gaze cues can affect fluency and efficiency of handover interactions between robots and humans were obtained.
- An exploration of how objects can be characterized by their affordances that makes it more comfortable for a human to receive objects being handed over has been conducted.
- A robot receiving system for automatically receiving an object from a human giver was developed.
- Experimental qualitative and quantitative results relating to changing parameters of robot receiving behaviours were obtained.
Chapter 2

Background

This chapter provides a brief overview on the current state-of-the-art work in research areas related to this thesis. In particular, it examines prior work directed towards enabling robots to participate in handovers through the use and recognition of nonverbal cues. Chapters 3 to 7, containing the main body of work within this thesis, will make reference to literature presented here. However, each of these chapters will contain additional background information that is more-specific to the topics covered by the chapter (e.g., machine learning, gaze cues).

Prior work has demonstrated that nonverbal cues play a significant role in coordinating handovers between human participants [43]. In particular, nonverbal cues provide an efficient and often unmindful modality of communication during human-human handovers that can help both giver and receiver fluently negotiate the transfer of an object in a matter of seconds. To obtain the same efficiency in human-robot handovers, researchers have examined human behaviours and cue generation that can lead to better robot interaction design. Studying nonverbal cues between humans has allowed some robot to mimic what people do when giving or receiving objects, leading to more time-efficient and desirable handovers [e.g., 22]. However, these observations are also particularly important for the development of a handover controller as robots may be able to anticipate rather than react to handovers, leading to more fluent interactions [50]. That is, if nonverbal cues indicating an object is being passed can be observed early enough, a robot may be
able to act quickly enough to allow the handover to proceed seamlessly and fluently from the perspective of a human giver [e.g., 106].

Previously, researchers have studied how some nonverbal cues in the spaces of proxemics, kinematics, dynamics and timing have affected the handover interaction in contexts of both human-human and human-robot dyads. This prior work is organized and presented with regards to these spaces and domains of nonverbal cues.

2.1 Proxemics

Much of the early work in robot-to-human handovers relied on proxemics studies by Hall, whose work suggested that distances between persons interacting socially are influenced by culture, attitudes, social standing, and relationships to one another [45]. In particular to human-human handovers, Basili et al. found that handovers occur approximately at half the distance between the giver and receiver and slightly to the right with respect to the receiver [13]. The interpersonal distances between givers and receivers varied considerably with the average distance being 1.16 m away roughly the distance where both giver and receiver must have arms outstretched during handover [13]. This spatial positioning is also corroborated by Huber et al. [53]. These studies did not examine factors leading to variances of interpersonal distances during handovers.

To compare this with proxemics between robots and humans, Walters et al. found that most of their participants allowed a robot to approach to within a personal distance (0.6 to 1.25 meters), suggesting that humans may treat robots similar to humans socially [121]. However, 40% of their participants allowed a robot to approach to within an intimate distance of 0.5 meters, implying that humans are more tolerant of robotic close encounters, which would be perceived as over-familiar or threatening in a similar human-human context. Similarly Koay et al. also found that participants would prefer a robot to approach to within approximately 0.6 - 0.75 meters in their study [67].
2.2 Kinematics

There have been numerous studies trying to better understand the kinematics involved in human-to-human handovers [e.g., 13, 50, 53, 60, 81, 104]. Several studies have looked at where handovers occur in the spatial domain [13, 54], and the joint/limb kinematics of both the giver and receiver during handover [41, 60, 106]. Kajikawa et al. and Koay et al. have both investigated motion planning of robots conducting handovers in close proximity to human counterparts, identifying typical kinematic characteristics in human-human handovers and developing appropriate handover trajectories based on such findings [60, 67]. Kajikawa et al. have determined that handovers between humans share several common kinematic characteristics, e.g., rapid increase in the giver’s arm velocity at the start of the handover, [60].

Other researchers have investigated robot handover trajectory and pose, reporting guidelines for how a robot arm should be positioned for handover and how that position should be achieved. For example, Agah and Tanie presented a handover motion controller that was able to compensate for unexpected movements of a human to achieve a safe interaction between human and robot [4]. Koay et al. identified human preferences for coordinated arm-base movement in the handover approach, observing that the majority of people preferred robots to approach a handover interaction from the front [67]. Pandey et al. and Mainprice et al. investigated the selection and recognition of handover locations based on the amount of human motion required to complete the handover [78, 97]. As a result of this work, Mainprice et al. designed an approach planner that considers both the mobility of the human receiver and robot giver in a cluttered environment. In this system, a mobile robot giver is able to plan the handover location and the approach to that location to allow for a handover that is safe, comfortable, and tailored to how mobile the human is. In a situation where the human receiver has low mobility (e.g., in a seated position, or is known to be infirmed), the robot plans a long path which gets it close to a human receiver to perform the handover, requiring little by the human. For a human receiver with high mobility (e.g., where the receiver is standing and/or known to be in good health), the robot plans a shorter path and chooses a handover location requiring shared effort by both the robot and human to reach that location.
In a related stream of work, Sisbot and Alami used kinematic features, along with preferences and gaze of the human receiver, to help a robot giver plan trajectories to navigate to a handover location safely and in a socially comfortable manner. In examining human-to-robot handovers, Edsinger and Kemp demonstrated study that humans inexperienced with robots were able to hand over and receive objects from a robot without explicit instructions. They also found that humans tended to control object position and orientation to match the configuration of the robot’s hand in order to make the robot’s task of grasping the object simpler.

Several works have examined how handover trajectories and final handover poses can best signal the intent to initiate a handover. Huber et al. used minimum jerk profiles in robot handover tasks to emulate human trajectories and behaviors. In their work, they found that participants in their experiment demonstrated a significantly shorter response time for minimum jerk profiles compared to trapezoidal joint trajectories, leading to a recommendation that minimum-jerk trajectory profiles be used for human-robot handovers. Glasauer et al. investigated how a robot can use human-like timing and spatial coordination of reaching gestures as an efficient and seamless method to convey the intent to handover an object and signal readiness. They note that their work, “shows that both the position and the kinematics of the [robot] partner’s movement are used to increase the confidence [of the human receiver] in predicting hand-over in time and space.” In building upon this work, Cakmak et al. and Strabala et al. found that the final handover pose should feature a nearly fully extended arm in a natural (human achievable) pose with the elbow, wrist, and distal point on the object positioned, respectively, from closest to furthest away from the body in all three dimensions. The object should be held in its default orientation and positioned to allow easy grasping by the human. Cakmak et al. defines default orientation as one in which an object is viewed most frequently in everyday environments. They note that “these orientations are often the most stable orientation for the object. Rotations around the vertical axis often do not effect the stability of the object, however for non-symmetric objects it can result in different functional properties.” In their study of five objects (plate, notebook, spice shaker, bottle, and mug), the default orientations were the upright orientation for the bottle, plate and shaker; upright positions with the handle on the right side for the mug; and the lying in a
readable orientation for the notebook. A related study emphasized the importance of the physical cues in human-robot handovers, showing that poorly designed handover poses and trajectories were often unsuccessful in communicating intent and ultimately resulted in handover failure [22]. In this study, the authors developed gestures through temporal (e.g., spatial variation of poses when transitioning into a handover) and spatial (e.g., difference in pose configurations of the robot arm amongst tasks) contrasts to provide fluidity and fluency to the handover interaction and in cueing when a user can successfully execute a handover [22]. They found that intent is best communicated by having high contrast between the pose used for holding the object and the pose used for handing over the object, which deterred users in reaching too early.

Recent studies investigating nonverbal gestures involved in handovers have examined more subtle kinematic cues such as gaze and eye contact. This work demonstrated that where participants look during handover plays a significant part in the coordination of handovers [83, 112, 113, 123]. A more comprehensive review of prior literature found on gaze cues can be found in Chapter 3 in Section 3.2.

2.3 Dynamics

In studies of grip forces during human-human object handovers, Mason and Mackenzie studied force profiles during transport and transfer of the object, finding that both giver and receiver use somatosensory feedforward control to synchronize transfer rate during handover [81]. Grip and load forces have also been shown by Chan et al. and Kim and Inooka to play a significant role in the coordination of handovers. Chan et al. showed that both the giver and receiver utilize similar strategies for controlling grip forces on the transferred object in response to changes in load forces. Through analysis of force loading on the transferred object, they found that the giver is primarily responsible for ensuring object safety in the handover and the receiver is responsible for maintaining the efficiency of the handover [26].
2.4 Timing

Kajikawa et al. observed the occurrence of a delay in handover reaches, as the receiver begins their reach for the object only after the giver achieves a maximum approach velocity [60]. Basili et al. quantified this delay by observing that nonverbal indicators suggestive of a handover interaction happened approximately 1.2 seconds prior to the occurrence of the actual handover [13]. In another stream of work by Admoni et al., it was found that delays in handover increased the amount of attention participants paid to the head, which increased human receivers awareness of nonverbal gaze cues [3]. The handover delay also increased compliance with those suggestions.

2.5 Summary

This chapter presented an overview of the literature on human-robot handovers that is relevant in the framing of this thesis. Previous studies demonstrate that nonverbal cues are a very necessary part of handover interactions, both between human and human/robot dyads. From this work, it is apparent that for human-robot handovers, robots are required to both emote and recognize nonverbal cues in a variety of different spaces in order to participate in fluent, efficient handover interactions. This thesis aims to build upon this previous work by continuing to explore how cues such as gaze and object orientation may affect handover interactions, both quantitatively and qualitatively. Also, in the latter half of this thesis, nonverbal cues in human-to-robot handovers are considered - a topic that has not been studied extensively in prior work.
Chapter 3

The Effect of Robot Gaze in Robot-to-Human Handovers

Aside from the shard recognition that the object is to transfer an object from the robot’s grasp into the human’s grasp without dropping the object during robot-to-human handovers, both robot givers and human receivers are additionally required to spatiotemporally coordinate their motions to achieve this objective. In human-human interactions, this coordination is done without verbal communication, but rather with subtle nonverbal cues. As seen in Chapter 2, prior work has shown that there are a great variety of subtle signals that may mediate human-human to create fluent and efficient interactions while ensuring the safety of the handover object (i.e., that it is not dropped). To investigate whether these subtle and natural signals occurring between human agents can improve robot-to-human handovers, these cues can be adapted for, and implemented on robot givers.

Using this exploratory framework, this chapter examines the first of the two research questions explored in this thesis: “How do nonverbal cues exhibited during robot giving and receiving behaviours change how users perceive the robot and affect the handover negotiation?” The studies presented in this chapter helps motivate the examination of nonverbal cues during handovers and provides an exemplar case study which focuses on the extent to which a robot’s use of nonverbal cues can profoundly impact the way that handovers are carried out between humans and robots as well as affect user experience of the task. More specifically, this work examines
how human-inspired robot gaze cues can affect spatiotemporal characteristics and subjective experience of a robot-to-human handover task.

In human-human social behaviour, gaze has shown to have a profound effect on communication as a signal for interpersonal attitudes and emotions. It is a cue medium that has shown to have vast bandwidth, being able to convey a wide variety of cues including attention, mood, attraction, intention, and specific conversational actions [8, 10, 61]. Because of its versatility, gaze has been studied within HRIS both as a method for a robot to better understand human agents [11, 111], and how anthropomorphic robots having a head can leverage gaze cues to establish more efficient communication with a human user [55, 90, 107]. In line with the latter stream of work, this chapter examines how certain gaze cues, inspired by observed gaze profiles used by humans during object handovers, play a role in establishing fluency and quality of a robot-to-human handover.

Studies presented in this chapter address whether gaze can be used to augment a robot-to-human handover by subtly communicating handover location and timing, while also providing an intuitive social interaction modality to help inform the human receiver’s behavioural decision on when and where to reach for the handover object. Gaze cues, in either human-human or human-robot interaction, have proven to be effective for communicating attention [66, 90]. During a handover, givers use verbal or nonverbal cues to direct the receiver’s attention to an object. Successful handovers typically take place when the two parties achieve shared attention on the same object. Previous studies [74, 112, 113] indicate that gaze can be used by robots to signal handover intent to users before the handover event. However, unlike these studies, the work presented in this chapter investigates the effect of robot gaze during the handover on the timing of the handover event. Here, gaze behaviours inspired via observations of human-human handovers were implemented on a PR2 humanoid robot. Through observation of user preferences and timing of the interaction, this research demonstrates that gaze can play a key role in improving the user experience of human-robot handovers, and help make handovers fast and fluent.

The contents of this chapter are largely transcribed from a manuscript of which I was one of the authors, entitled, “Meet me where I’m gazing: how shared attention gaze affects human-robot handover timing” [88]. This manuscript was sub-
mitted to the 9th Association for Computing Machinery (ACM)/IEEE International Conference on Human-Robot Interaction held from March 3-6, 2014 at Bielefeld University in Bielefeld, Germany. This work was awarded “Best Paper” at the conference. The version of the manuscript that appears in this chapter contains several minor modifications to the original conference paper. These modifications include:

- Figures have been updated for enhanced clarity.
- The introduction contained in Section 3.1 has been abridged.
- Background information originally found in Section 3.2 that reiterates the literature review contained in Chapter 2 has been removed.
- Terminology has been updated to match terms used in this thesis.
- Experimental conditions have been renamed for clarity.
- Discussion (Section 3.7) and conclusions (Section 3.8) of this work have been expanded to address impact and limitations of the work presented in the chapter, as well as directions for future work.

Supporting materials for this work including advertisements, consent forms and surveys can be found in Appendix A.

### 3.1 Introduction

This work seeks to improve human-robot handovers by investigating how gaze can be used to augment a handover event, subtly communicating handover location, handover timing, and providing acceptable social interaction signals. Gaze cues, in either Human-Human Interaction (HHI) or in HRI, have proven to be efficient for communicating attention [66, 90]. During a handover, givers use verbal or nonverbal cues to direct the receiver’s attention to an object. Successful handovers typically take place when the two parties achieve shared attention on the same object. Previous studies indicate that gaze can be used by robots to signal handover intent to users prior to the handover event [74, 112, 113]. However, these studies did not explore the effect of robot gaze during the handover on the timing of the handover event.
It is hypothesized that the use of gaze cues during human-robot handover can influence handover timing and the subjective experience of the handover by implicitly increasing communication transparency and perception of naturalness for the interaction.

In this work, two studies are conducted: a human-human study (Section 3.3) followed by a human-robot study (Section 3.4). The human-human study was conducted to observe gaze patterns used during human-human handovers. The results of this first study were used to inform the second study on robot gaze during robot-to-human handovers: in this latter human-robot study, the two most frequently observed gaze patterns from the first human-human study were implemented along with a ‘No Gaze’ condition on a PR2 humanoid robot platform (Willow Garage Inc., Menlo Park, California, USA) in the giver role. The following two questions are addressed through these studies:

1. Can gaze improve the subjective experience of handovers?
2. Can gaze be used to produce faster, more fluent handovers?

The key results from this work are as follows:

- Subjects reach for the object significantly faster when the robot directs its gaze towards the intended handover location (Shared Attention gaze) than when no gaze cues are used.
- Subjects tend to perceive a handover as more natural, communicative of timing, and preferable when the robot provides Turn-Taking gaze in addition to Shared Attention gaze.

### 3.2 Background - Gaze in Handovers

Gaze is an important and useful cue in HHI. People repeatedly look at each other in the eye during social interaction and people do not feel that they are fully engaged in communication without eye contact [9]. Studies in psychology have shown various functions of gaze in social interaction, such as seeking and providing information, regulating interaction, expressing intimacy, exercising social
control, etc. [9, 66, 98]. Gaze can be named differently in different social situations [90]; for example, mutual gaze or eye contact is defined as two people looking into each other’s face or eye region [119], while deictic gaze or shared visual attention is defined as one person following the other’s direction of attention to look at a fixed point in space [20].

Previous work has shown the importance of gaze in HRI. For example, Staudte and Crocker [111] demonstrated that humans react to robot gaze in a manner typical of HHI. Since gaze behaviour is closely linked with speech [8], much work has been done on the conversational functions of gaze in HRI [72, 75, 83, 89, 91, 118]. Gaze is particularly effective in regulating turn-taking during human-robot conversation. Kuno et al. [72] developed gaze cues for a museum guide robot to coordinate conversational turn-taking. Matsusaka et al. [83] used gaze cues to mediate turn-taking between participants in a group conversation.

Another large body of literature focuses on using gaze to direct people’s attention in HRI [16, 47, 56, 100, 105]. Gaze was combined with pointing gestures in [16, 47, 56] to direct people’s attention, which the authors believed would make the interaction more human-like [16] while minimizing misunderstanding [47]. In Rich et al.’s work, four types of connection events were identified from HHI videos, namely directed gaze, mutual facial gaze, adjacency pairs and back-channels. Implementing them in an HRI game showed a high success rate in forming human-robot connection or joint attention [100]. In work done by Sidner et al., people directed their attention to the robot more often in interactions where gaze was present, and people found interactions more appropriate when gaze was present [105].

Introducing gaze cues can also benefit HRI in other ways. Mutlu et al. and Skantze et al. showed that gaze increased human performance in certain human-robot tasks [89, 107]. Other studies have shown that gaze heightened human-robot engagement and contributed to the perceived naturalness of a communicating robot [72, 75, 105].

In the study of human-robot handovers, other researchers have shown that gaze can be useful in communicating the intent to initiate a handover. Lee et al. [74] studied human motion and gaze cues as people approached each other for handovers. They found that people looked at the object or at the receiver as they ap-
proached the receiver. Strabala et al. [112] examined the signals that humans use to communicate handover intent before a handover takes place. They initially acknowledged gaze as one of the important features that mark the difference between different phases in handover, but they did not find gaze to be an effective predictor of handover intent. In contrast, Kirchner et al. [65] demonstrated how robot gaze can be effective in targeting an individual recipient out of a group of people for a robot initiated handover. Atienza and Zelinsky [11] augmented handover interactions with gaze, demonstrating a system that allowed a human to request an object for handover by looking at it.

While the above studies addressed gaze in pre-handover cuing and communication of intent to handover, this work examines the use of gaze during the handover event. Although the effectiveness of gaze in regulating handover intent remains an open question, gaze may have a positive effect when used during the handover event, since it helps establish shared attention and has been shown to improve human-robot tasks. Hence, it is hypothesized that gaze may be useful in improving the handover itself by establishing shared attention and signalling the robot’s end of turn.

3.3 Study I: Observing Gaze Patterns in Human-to-Human Handovers

To inform the human-robot handover study design, a study on gaze behaviours of human-to-human handovers was conducted. In this study, a bottle of water was handed over between participants acting as giver and receiver multiple times. The gaze behaviour of the giver was recorded and observed during the handover. While other researchers have observed gaze in human-human handovers before the handover event, [e.g., 112], the work described here aims to augment these previous results with a study focusing on gaze during the handover event.

3.3.1 Experimental Procedure

Twelve volunteers (10 male, 2 female) participated in this study in pairs. The giver was asked to handover ten bottles from a side table to the receiver one at a time. The receiver was asked to bring the bottles to a collection box about two
meters behind them one at a time, requiring him/her to walk away from the giver between each handover. This process repeated until all ten handovers were completed. Each participant performed the role of the giver, then was paired with another participant and performed the role of the receiver, resulting in a total of twelve giver-receiver pairs (120 handover events in total).

In order to collect human gaze patterns that can inform the design of human-robot handovers, the giver and receiver were instructed not to talk during this process. The giver was also instructed to pick up the bottles from the side table only after the receiver returned from the collection box and had put his/her hands on the table. By requiring the receiver to turn and walk away, the common attention between the giver and receiver was interrupted after each handover and participants needed to re-connect for the next handover.

3.3.2 Results

Annotation of a frame-by-frame video analysis of the givers’ gaze patterns indicates that the giver’s gaze during a handover can shift between three positions: the object being transferred, the expected handover position, or the receiver’s face. Through a frame-by-frame analysis, givers’ gaze patterns from video recordings of the 120 handovers were annotated. From this process, the following gaze patterns were found (Figure 3.1):
Figure 3.1: Giver’s gaze patterns observed from human-human handovers. ‘Shared Attention’: continual shared attention gaze; ‘Face’: continual face gaze; ‘Turn-Taking’: long Shared Attention gaze followed by a short Face gaze; ‘Short Face Shared Attention’: short Face gaze followed by a long Shared Attention gaze; ‘Long Face Shared Attention’: long Face gaze followed by a short Shared Attention gaze.
• **Shared Attention Gaze** (Figure 3.2 Top): The most frequent gaze pattern (68% of all handovers observed) consists of the giver gazing at a projected handover location as s/he reaches out to execute the handover. After picking up the bottle, the giver turns to face the receiver, looks down at a midpoint between the giver and the receiver, and keeps the gaze there until the receiver takes control of the bottle. This midpoint is approximately where the handover takes place. There is no eye contact between the giver and the receiver throughout this handover gaze pattern.

• **Face Gaze**: In some other (10%) handovers, the giver gazes at the receiver’s face, perhaps to establish an eye contact, throughout the handover. This gaze behaviour towards the receiver’s face is labeled ‘Face Gaze’.

• **Turn-Taking Gaze** (Figure 3.2 Bottom): In 9% of handovers, a slight variation from the Shared Attention Gaze was observed. In addition to gazing at a projected handover location while reaching out, the giver also looked up to make eye contact with the receiver near the end of the handover motion, at approximately the time that the receiver made contact with the bottle.

• **Short Face Shared Attention Gaze**: In 8% of the handovers, the giver looked at the receiver’s face and quickly glanced at the bottle when the receiver is about to touch the bottle.

• **Long Face Shared Attention Gaze**: In 5% of the cases, the giver glanced at the receiver before but not during handover, and shifted the gaze to the handover location when the receiver is about to touch the bottle.

### 3.3.3 Discussion

The results of this study suggest that humans use a variety of gaze patterns while handing over an object to another person. In general, the giver tends to shift his/her gaze from the object being handed over to the receiver’s face (Face gaze), the projected location at which the handover should take place (shared attention gaze), or a combination of the two. The shared attention gaze in the Shared Attention, Turn-Taking, Short Face Shared Attention, and Long Face Shared Attention
Figure 3.2: Demonstration of two frequently observed gaze behaviours from the human-human handover study. Top: *Shared Attention* gaze - the giver looks at the location where the handover will occur. Bottom: *Turn-Taking* gaze - the giver looks up at the receiver after the shared attention gaze.
patterns can be interpreted as serving the function of communicating where the physical transfer of the object should happen. The long Face gaze in the Face and Long Face Shared Attention gaze patterns serves a similar function as the face gaze in verbal conversation of regulating a turn [65]; in verbal conversations, the speaker typically ends his/her utterance with a sustained gaze at the listener, signaling willingness to hand over the speaker role, while in this case, to hand over the object. The short Face gaze in the Turn-Taking and Short Face Shared Attention gaze patterns serves a monitoring function [65], appearing to observe whether the receiver is paying attention to, or is ready for, the transfer of the object. These results inspire the hypothesis that an implementation of analogous gaze cues for a human-robot handover could serve similar functions and help the human-robot dyad perform more fluently. Hence, the gaze patterns observed in this study informed the design of experimental conditions in a second study. Section 3.7 provides more discussions of the results from the two studies contrasting human-human and human-robot handovers.

3.4 Study II: Impact of Human-Inspired Gaze Cues on First-Time Robot-to-Human Handovers

To examine the impact of robot gaze on human receiver behaviour, a PR2 humanoid mobile robotic platform (Willow Garage Inc., Menlo Park, California, USA) was used with a pan-tilt head and two 7-DOF arms, each with a two-fingered, 1-DOF gripper. In the following section, the physical handover cues used by the PR2 are outlined (Section 3.4.1), describe the three experimental gaze conditions inspired from the human-human handover study (Section 3.4.2), and outline the experiment design and technical implementation (Sections 3.4.3 and 3.5).

3.4.1 Physical Handover Cues

In designing the robot’s motions performed during handover, we looked towards work done by Cakmak et al. and Strabala et al., which emphasized the importance of the physical cues in human-robot handovers, showing that poorly designed handover poses and trajectories were often unsuccessful in communicating intent and ultimately resulted in handover failure [22, 113]. They found that intent
is best communicated by having high contrast between the pose used for holding the object and the pose used for handing over the object. We have followed these guidelines in the design of our handover poses and trajectories.

In the experiment, based on the findings of Basili et al. [13] and Koay et al. [67], the robot was positioned such that it was facing the participant approximately 1 meter away. The robot executed the handover with its right gripper, as recommended by Koay et al.. In the beginning of each handover, the robot starts its motion at the *Grasp Position* with its end-effector prepared to grasp a bottle sitting on a table at the robot’s right side. When subject is ready, the end-effector grabs the bottle (marking a start time, $t = 0$ of the interaction), then moves the bottle horizontally to a position in front of the robot’s center-line (*Ready Position*). Then the robot moves from the *Ready Position* forward to the handover location. The joint-angle goals of the *Grasp Position*, *Ready Position*, and *Handover Location* were predefined such that when the robot’s end-effector is extended, the arm is positioned in a natural pose: the elbow located below the shoulder, and the gripper located below the distal point on the bottle, as shown in Figure 3.3. The *Handover Location* was selected in accordance with the recommendations of previous work [65, 74]. The three locations are constant for all three gaze conditions. While other researchers have proposed handover controllers that adapt to the position of the human’s hand, [e.g., 37], a constant *Handover Location* was maintained throughout the experiment and only vary gaze cues used during handovers.

When the robot’s arm reaches the *Handover Location*, the robot waits for a participant to grasp and pull up on the object. The force the gripper exerts on the bottle is a linear function of the downward force exerted by the bottle as described by Chan et al. [27]. Thus, as the subject takes the weight of the bottle, the robot releases its grip (marked as the release time). The PR2’s fingertip pressure sensor arrays were used to realize Chan et al.’s handover controller. Finally, after releasing the object, the robot returns to the *Grasp Position*, ready to grasp and lift the next object.
Figure 3.3: Illustration demonstrating the experimental set-up and the three conditions at the Handover Location: a) No Gaze; b) Shared Attention; and c) Turn-Taking. An array of infrared sensors was located at the edge of the table. The green dotted lines represent the location where subject’s reach motion is detected. Participants stood at a specified location marked on the floor.
3.4.2 Experimental Gaze Cues

In this study, the PR2 robot expressed gaze through head orientation. Imai et al. [55] showed that robot head orientation can be an effective substitute for human-like gaze and that head orientation is interpreted as gaze direction. In order to minimize any possible confusion regarding the robot’s gaze direction, a single object for the handovers was used.

Three different gaze patterns in human-robot handovers were tested, as shown in Figure 3.4. In all conditions, the robot’s gaze tracks its end-effector from the Grasp Position to the Ready Position as though the robot is attending to the acquisition of the bottle. When the end-effector arrives at the Ready Position, the robot’s head is tilted downwards towards the end-effector. Only when the robot arm transitions between the Ready Position to Handover Location does the robot transfer its gaze according to the following gaze patterns.

The No Gaze gaze condition acts as the baseline for this study. The robot head remains looking down towards the ground while the end-effector extends forward for the handover.

The Shared Attention gaze condition models the most frequently observed gaze pattern from the human-human handover study. When the robot starts to move from the Ready Position to the Handover Location, it smoothly transitions its gaze (head orientation) from the bottle to the location in space where the handover will occur, as an implicit cue intended to direct the human’s attention towards the projected Handover Location. With this condition, the hypothesis that shared attention can be established through gaze during handovers is testable, and that doing so benefits the handover interaction. Establishing shared gaze at an object or location can serve to direct shared attention (e.g., [56]) and can aid in the successful execution of human-robot cooperative tasks (e.g., [107]).

The Turn-Taking gaze condition is also derived from the human-human handover study, and is analogous to the third most frequently observed gaze pattern. This gaze condition was implemented on the robot over the second most frequently observed gaze pattern - the Face gaze - as it appeared to offer more information to the user i.e., a turn-taking timing cue, and would pose as a more interesting experimental condition. When the handover trajectory begins, the robot smoothly
Figure 3.4: Diagram depicting the gaze cues for the No Gaze (bottom), Shared Attention (middle), and Turn-Taking (top) gazes. In the Turn-Taking condition, the robot shifts its gaze from the Handover Location to the human’s face midway through the handover motion.
transfers its gaze to the Handover Location, as in the Shared Attention condition, but then shifts its gaze up to the human’s face in a quick motion, reaching the final gaze position at approximately the same time that the handover motion completes. Here, two hypotheses are tested: that (a) this gaze shift can cue handover timing, and (b) looking at the face can improve the subjective experience of the handover. This type of gaze shift has been shown to be a meaningful human-robot turn-taking cue [23] and mutual gaze can increase the sense of engagement and naturalness in human-robot interactions [75,105].

3.4.3 Experimental Procedure

A paired-comparison handover study was conducted in a controlled room. The study took place on the day of a university orientation event such that many and diverse naive participants could be rapidly recruited during the public event. Section A.1 shows the advertisements used for recruitment. The experiment was structured as a balanced incomplete block design \( (v = 3, b = 96, r = 64, k = 2, \lambda = 32) \)\(^1\) to both support rapid trials (maximum 5 minutes) and include only naive reactions: each participant evaluated one of the three condition pairings. Condition order was randomized and presentation order counterbalanced among trials.

Participants read a consent document, shown in Section A.2.1, outlining the experimental procedure and risks. Each participant provided verbal informed consent (Section A.2.2), then entered the room where instructions were given. They were told to stand at a marked area facing the robot, and informed they would participate in a handover interaction. Participants were also told that the robot would pick up the water bottle placed beside it and hand it to them. They were asked to take the bottle from the robot whenever they felt it was the right time to do so. To avoid unintended cuing, during handovers the experimenters sat out of the field of view of participants.

After receiving the first bottle, participants placed the bottle in a box approximately 3 meters behind him/her. This served as a washout between handovers.

\(^{1}\)These variables indicate the structure of a balanced incomplete block design: \( v \) = number of treatments (i.e., three conditions were tested: No Gaze, Shared Attention, and Turn-Taking gaze conditions); \( b \) = number of blocks (i.e., observations from a total of 96 participants was analyzed); \( r \) = number of replicates (i.e., each condition was tested on a total of 64 participants); \( k \) = block size (i.e., each participant saw two conditions); and \( \lambda = r(k - 1)/(v - 1) \).
breaking the participant’s focus on the robot and the handover, as was done previously by Cakmak et al. [22]. Participants then filled out a short questionnaire regarding the handover interaction they just conducted (shown in Section A.3.1), returned to the same marked area front of the robot and participated in a second handover. Participants were permitted to keep the last bottle given to them by the robot.

During each handover, the following events were timestamped: start of robot motion (start time), end of robot motion (end of motion time), start of release of the robot’s gripper (release time), and the participant’s first reach for the object (reach time) as measured by the motion sensor array described in Section 3.5.

After the second handover, participants left the room and completed a short questionnaire (shown in Section A.3.2) comparing the two handovers on three subjective metrics: overall preference, naturalness, and timing communication. For each of the following three questions, participants were asked to select either the first or second handover:

1. Which handover did you like better, overall?
2. Which handover seemed more natural?
3. We think that timing is important in human-robot handovers. Which handover made it easier to tell when, exactly, the robot wanted you to take the object?

Participants could also provide additional comments.

3.5 Technical Implementation

To control the PR2, the Robot Operating System (ROS) (Willow Garage Inc., Menlo Park, California, USA) used by the PR2 was extended with a series of software modules coordinated via the Blackboard architectural design pattern [48] as shown in Figure 3.5. One module controlled the robot’s arm and another, its head. The head-control module provided object tracking functionality for bringing the object to the Ready Position, and a smooth, fast gaze transition (average 90 degrees/second) functionality to enable the Shared Attention and Turn-Taking gaze
conditions during the handover motion. An independent module logged quantitative measurements of robot’s start time, end of motion time, and release time.

An array of three passive SEN-08630 infrared motion sensors (SparkFun Electronics, Boulder, Colorado, USA) configured as a light curtain was placed at the edge of the table (as shown in Figure 3.4), and was used to detect the start of the participant’s reach (reach time) triggered by the participant’s hand crossing the table edge. An Arduino UNO microprocessor (Arduino LLC, Turin, Italy) relayed the sensor reading to the computer controlling the robot. Sensor readings were logged and time-synchronized with the robot.

3.6 Results

A total of 102 volunteers participated in the experiment. Six records were rejected as instructions outlined in Section 3.4.3 were not followed. Subsequently, data from 96 participants (33 females, 63 males), aged 18-61 years \(M=23, SD=5.59\) was analyzed. Due to technical error, reach time was not logged in the second handover for five of the participants. This did not affect the analysis of handover timing, since the focus of this study the focus of this study on first-time responses requires reach time measures from only the first handovers. No other technical failures occurred and all handovers were successful, i.e., no bottles were dropped.

3.6.1 Handover Timing

Figure 3.6 shows the distribution of three key times: the robot’s end of motion time, participant’s reach time, and robot’s gripper release time. All times are measured relative to the start time of the interaction - i.e., the robot’s start of motion. A comparison using unpaired t tests of these times to human-human handovers observed by Basili et al. show that duration between handover initiation to the giver’s release of the object is significantly slower (by approximately two seconds) for human-robot handovers compared to the human-human handovers \(p < .0001\) [13].
Figure 3.5: System flow diagram describing algorithm used for PR2 robot gaze during handovers.
Figure 3.6: Chart of handover timing results for various gaze conditions in the robot-to-human handover study. All times are measured with respect to the robot’s start of motion at $t = 0$. The dashed line at 2 seconds indicates the end of robot motion at the Handover Location. Reach time indicates the participant’s reach toward the proffered object crossing the infrared sensors. Note that in the case of the Shared Attention condition, participants start to reach before the robot has reached Handover Location. The mean reach time for the Shared Attention condition is significantly earlier than that of the No Gaze condition. Error bars indicate 95% confidence intervals.
A one-way Analysis of Variance (ANOVA) was conducted on participants’ reach time across the three conditions. A significant learning effect between the first and second handover trials is observed \[ t(90)=4.21, \quad p<.001, \quad d=0.43 \]. In this learning effect, reach time is significantly earlier for the second set of handovers for participants. As the goal of this work is to understand first time, inexperienced behaviour, only the reach time collected during the first of the two handovers performed by each participant was used. The entire robot motion from the Grasp Position to the Handover Location consistently took 2.02 seconds (\( SD = 0.01 \)).

Participants’ reach time varied across the three gaze conditions \( F(2,93)=6.49, \quad p<.005 \) as plotted in Figure 3.6; post-hoc analyses used a Bonferroni correction. Participants reached for the object significantly earlier with Shared Attention \( [M=1.91, \quad SD=0.52] \) than with No Gaze \( [M=2.54, \quad SD=0.76] \) (\( p < .005 \)). Note that the mean reach time for Shared Attention occurs before the robot has stopped moving at the Handover Location (reach time < end of motion time). No significant differences were found between Shared Attention and Turn-Taking \( [M=2.26, \quad SD=0.79] \), or between Turn-Taking and No Gaze.

### 3.6.2 Subjective Experience

Durbin’s test [35] was employed to contrast overall preference, perceived naturalness, and timing communication across the three gaze patterns during handovers on the questionnaire data. This test is analogous to a Friedman test for rank data, but adapted to balanced incomplete block designs.

Possible gender effects were checked using Mann-Whitney U tests. No significant effects of gender were found (overall preference: \( U=935.0, \quad p=.23, \quad r=0.12 \); naturalness: \( U=918.5, \quad p=.18, \quad r=0.14 \); timing communication: \( U=935.5, \quad p=.22, \quad r=0.12 \)). One-sample Wilcoxon signed rank tests allowed for observation of potential bias in selecting the first or second handover experience in the questionnaire. From these tests, significant biases towards selecting the second handover on the timing communication metric \( Z=2.22, \quad p<.05 \) and a weak trend to select the second handover on both overall preference \( Z=1.62, \quad p=0.11 \) and naturalness metrics \( Z=1.41, \quad p=0.16 \) were found. The rank data collected using the questionnaire is insufficient to correct for this bias statistically.
Given this general bias to select the second handover, finding statistical significance to $\alpha = 0.10$ in questionnaire results is noteworthy \cite{116}. Hence, observation of trends having $p < 0.10$ is reported in Table 3.1.

**Table 3.1:** Ranking of questionnaire results. Each cell represents the number of people who chose the row condition over the column condition. * indicate pairwise comparisons that are significant to $p < 0.10$ (none were significant to $p < 0.05$). Note that participants’ bias to select the second handover experience regardless of experiment condition was observed to be significant.

<table>
<thead>
<tr>
<th>Overall Preference</th>
<th>Condition</th>
<th>Turn-Taking</th>
<th>Shared Attention</th>
<th>No Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn-Taking</td>
<td>-</td>
<td>21*</td>
<td>19*</td>
<td></td>
</tr>
<tr>
<td>Shared Attention</td>
<td>11</td>
<td>-</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>No Gaze</td>
<td>13</td>
<td>15</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Naturalness</th>
<th>Condition</th>
<th>Turn-Taking</th>
<th>Shared Attention</th>
<th>No Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn-Taking</td>
<td>-</td>
<td>20*</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Shared Attention</td>
<td>12</td>
<td>-</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>No Gaze</td>
<td>13</td>
<td>13</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Timing Communication</th>
<th>Condition</th>
<th>Turn-Taking</th>
<th>Shared Attention</th>
<th>No Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn-Taking</td>
<td>-</td>
<td>21*</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Shared Attention</td>
<td>11</td>
<td>-</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>No Gaze</td>
<td>14</td>
<td>13</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Overall Preference:** No significant difference in user preference across the three gaze conditions was found [$T_2 = 2.04, p = 0.14$]. However, one-tailed pairwise comparisons demonstrate a trend for preference toward Turn-Taking over No Gaze ($p < 0.10$) and Shared Attention ($p < 0.10$) conditions.

**Naturalness:** No significant difference in perceived naturalness of the handovers across the three gaze conditions were found [$T_2 = 1.82, p = 0.17$]. How-
ever, participants tended to choose Turn-Taking as more natural than Shared Attention ($p < .10$) but not over the No Gaze condition.

**Timing Communication:** No significant differences were found in the perceived communication of timing across the gaze conditions [$T^2 = 1.65, p = .2$]. However, participants tended to choose Turn-Taking over Shared Attention ($p < 0.10$), but not over No Gaze, as easiest to communicate handover timing.

In total, 59% of all participants provided additional comments (optional) on the questionnaire. Twelve subjects who experienced the Turn-Taking condition explicitly used words such as “head motion”, “eye contact” or “looking at me” and expressed the condition in a positive light (e.g., Participant 90 compared No Gaze with Turn-Taking: “During second handover [Turn-Taking], robot made eye contact, which made it easier to tell when the bottle should be taken.”; Participant 10 compared Shared Attention and Turn-Taking: “I liked it when robot looked at me. That confirms it’s good to take.”). However, another twelve subjects expressed that they did not notice any difference between the conditions.

### 3.7 Discussion

Building on previous work that studied communication of intent to handover using gaze, these two studies explored how gaze during handovers (i.e., after the intent to handover has already been communicated and while the handover is taking place).

A handover interaction typically involves the well-defined role assignments of a giver and receiver, and a clear sequence of actions that must take place - i.e., the giver grabs and transports the object, and then the receiver receives the object from the giver. However, even in such a well-defined interaction between two agents, the details of when and where must be dynamically negotiated for the interaction to be successful.

As an effort to understand what nonverbal cues may help in this process, the first study (Study I) was conducted to identify gaze patterns human givers use when handing over an object to another person. Results of this study not only provide five different gaze patterns human givers use in an human-human handover, but it
also suggests that the observed patterns more often than not involve gazing at the projected handover location.

Following Study I, Study II explored the impact of robot gaze on human-robot handover timing and user perception of the handover experience. Results show that participants reached for the proffered object significantly earlier when the robot performed a Shared Attention gaze at the projected Handover Location. In fact, participants reached for the object even before the robot had arrived and stopped at the Handover Location (a mean of 0.11 seconds before the end of motion time). This is in contrast to the No Gaze condition where the mean reach time is 0.52 seconds after the robot’s end of motion. In this study, participants were explicitly told that the robot would be handing over objects to them and that they were to take the object from the robot. In addition to this foreknowledge, the robot used highly contrasting poses between the Ready Position and the Handover Location which, according to Cakmak et al. [22], makes the robot’s intent to hand over the bottle very clear. Hence, it is unlikely that the observed difference in timing between the gaze conditions is due to uncertainties in understanding the robot’s handover intent. Rather, the results suggest that the robot’s gaze at the projected Handover Location supplements the communicated intent with implicit information on where the handover should take place. This may be helping to establish shared attention on the Handover Location even before the robot arrives there, naturally allowing participants to respond and meet the robot at the location earlier than when such a cue is absent. Thus, the result best supports an increase of fluidity in the execution of the handover as it takes place.

However, the role of mutual gaze used in the Turn-Taking condition required further investigation. At the beginning of the robot’s handover motion, the robot expresses the same locational, shared attention gaze in both the Shared Attention and Turn-Taking conditions. Hence, it was surprising to find that the reach time of the Turn-Taking condition is not significantly earlier than that of the No Gaze condition.

The Turn-Taking condition was intended to test two hypotheses: that the Turn-Taking gaze would cue handover timing, and that looking at the participant’s face would improve the subjective experience of handover. While a trend was observed suggesting that the robot’s gaze directed at the face improves the subjective ex-
perience of handover, it appears that the shared attention gaze cues the handover timing instead of the mutual, turn-taking gaze. It may be that the mutual gaze implemented for this study served the function of acknowledgement rather than the intended function of turn-taking. Qualitative differences in participants’ reaction may exist between the Shared Attention and Turn-Taking conditions. For example, participants may have started to respond to the robot’s shared attention gaze in both conditions but, prior to being detected by the reach sensor, saw the robot’s turn-taking gaze and stopped to make eye contact before continuing to reach across the table.

This raises unexplored questions about how participant’s reach time is affected by the timing of the robot’s gaze. How much would varying the robot’s gaze timing affect human reach time? Is the timing of the robot’s gaze a more dominant cue than the location the robot is gazing at? That is, would a robot that shifts its gaze from the object directly to the person’s face during handovers have the same effect as the Shared Attention condition? Would we see changes in participants’ reach direction if the robot gazed at a different location? Without a thorough qualitative analysis, it is difficult to tell, with accuracy, if and when shared attention is established with the participant. A separate experiment with a gaze tracking device would help answer these questions, and is left as future work.

It is important to note that the results may be representative of first-time, inexperienced participants only, where novelty effects may have motivated them to observe the robot more carefully than they would if they were more familiar with the robot. Unsurprisingly, a significant training effect was observed in the reach time data, as well as a bias toward describing the second handover experience more favourably in the questionnaire regardless of the condition experienced. Some of the participants’ comments suggest that in certain cases, people did not pay attention to the head of the robot at all. Indeed, it is suspected that in many human-human handover scenarios, especially those that are repetitive or trained (e.g., passing a baton in a sprint relay race), people do not use gaze cues at all and yet succeed in object handover. Thus, it is hypothesized that robot gaze cues may not have the same effect on trained or familiarized users.

Although the earlier reach time of participants in a handover may seem more similar to natural, unscripted human-human handovers, this may not necessarily
be desired in some human-robot handover situations. Depending on the handover controller implemented on a robot, handover timing may need to be controlled such that people naturally grab the object only when it is safe to do so. Many of the handover controllers that modulate the release time of the object are built for cases where the robot’s gripper is already at the Handover Location before people grab the object. A situation where the object is grabbed before the robot is ready to release the object could lead people to pull hard on the object, possibly damaging or dropping the object, or resulting in a negative perception of the robot.

In the context of this thesis, this work motivates the examination of nonverbal cues within human-robot handovers and shows that the use and adjustment of robot gaze profiles can significantly alter the timing and user experience of the interaction. Although we find that handovers between robots and humans observed in this study are still significantly slower than handovers between humans, the significant difference in interaction duration between the Shared Attention and No Gaze gaze conditions supports the notion that the appropriate design of robot gaze profiles may lead to a more streamlined encounter. Beyond gaze, the finding of these studies presented here suggests that carefully designing nonverbal cues for robots, perhaps through observation of human behaviours, can indeed improve fluency and efficiency of human-robot handovers.

### 3.8 Conclusions

Previous work on gaze in human-robot interactions has offered evidence that gaze can affect human behaviours. The studies presented in this chapter examine the effect of robot-generated gaze cues can effectively influence details of human action for a successful, fluent robot-to-human handover. Results from a human-human handover study identified five types of gaze patterns that humans tend to use during handovers. In a second study observing robot-to-human handovers, a robot mimicked two gaze cues identified from the first study (Shared Attention and Turn-Taking), showing how a robot’s use of human-inspired gaze expressions can affect the timing of a robot-to-human handover in first-time participant responses. The study provides empirical evidence that a human-inspired gaze pattern (Shared Attention) implemented on a robot can elicit a human receiver to reach for and
retrieve the proffered object earlier compared to a condition where no gaze cues (No Gaze) are offered.

Addressing the first research question of this thesis, this work has demonstrated how a robot’s human-like use of gaze expressions in human-robot handover can affect the timing of the handover event. Once the intention to handover is established, providing a gaze cue to the projected Handover Location seems to offer rich gestural information about the handover event, and allows users to reach for the object sooner, even before the robot arrives at the location.

3.8.1 Limitations

Cross-culturally, gaze may signal differently to people of various ethnic origins, and thus the results of this study may not be reproducible outside of North America. As the results were obtained from a majority sample of non-expert users having little to no experience with robotics that were interacting with the robot in a controlled setting, the findings may not applicable to users who will have had extensive opportunity to interact with robots in the field. As the results of the human-robot study show the occurrence of a learning effect happening with participants, we would expect findings will differ for expert users.

3.8.2 Future Work

As a result of this work, there are several directions for future inquiry that may examine research questions related to gaze in handovers. One avenue of possible work examines direction of gaze: if the robot’s gaze was intentionally directed to a different location rather than the predicted handover location or the receiver’s face (perhaps in an attempt to intentionally mislead the user) would there be changes to the handover timing or participants’ reach direction? If so, it would offer additional evidence to the importance of gaze as a nonverbal cue that may be important for other HRIS. Another research question that would be interesting to explore relates to if, when and under what conditions shared attention mutual gaze is established between robot and human during a handover event. Furthermore, do users prefer eye-contact gaze in addition to the locational, shared attention gaze? An experi-
mental setup with a gaze tracking device would help answer these questions in a future study.
Chapter 4

Characterization of Handover Object Orientations Informing Efficient Robot-to-Human Handovers

The work presented in the previous chapter showed that subtle gaze cues generated by a robot in a robot-to-human handover can have a significant impact on the efficiency and fluency of the handover. In reference to the first research question posed in Section 1.2 observing that gaze cues can affect these aspects of an interaction suggests that other nonverbal cues may have significant, and perhaps unique, consequences within the scope of HRIS. This chapter continues to investigate cues for robot-to-human handovers, though begins to address the second research question of this thesis: how can a robot adequately recognize and interpret nonverbal cues conveyed by a human giver to infer object attributes? In particular, focus is given to a more implicit cue: the orientation of objects as they are being passed from giver to receiver. From observation of these orientations, the aim of this work is to determine if and how a robot may be able to infer an object’s affordances.

Objects’ affordances and how they are oriented during handovers have a significant impact on the handover experience, and in particular, for the receiver of
a handover object [5]. Good handover orientations - those that direct object affordances such as a handle towards the receiver - assist in maintaining handover efficiency and fluency. On the other hand, bad handover orientations which do not consider object affordances may not only handicap efficiency and fluency, but may also cause awkwardness, mistaken intentions and/or injurious results. Figure 4.1 illustrates examples of both good and bad handover orientations. To ensure that robots act as well-behaved givers, they must be trained to consider appropriate object orientations during handovers. Specifically, robots must always direct the object’s affordances towards the receiver for a more fluent handover. This chapter will examine natural handover orientations used by people as a step towards enabling robots to learn how to appropriately orient objects during robot-to-human handovers.

Figure 4.1: Illustration depicting the importance of considering object orientation during robot-to-human handovers. On the left, the orientation of the knife used by the robot during the handover allows the user to grasp the knife’s affordance, the handle, easily. On the right, the robot does not consider the object’s affordance when attempting to perform a handover; thus, the intent of the robot is not fluently communicated, the interaction is awkward/dangerous, and the human no longer wishes to participate in future HRI experiments.

In this work, a user study is performed to examine orientations of 20 common, household objects during handovers between people. This study compares how unprompted humans naturally orient handover objects to orientations that are giver-
and receiver-centered in terms of the objects. The intent of this comparison is to examine if observations of natural handovers between humans can be used as templates by robot givers to achieve more efficient and socially acceptable handovers; that is, can robots assume that givers in natural handovers use object orientations that are most considerate to the receiver? Through this exploration, this work also introduces the notion of affordance axes for certain objects as a method of comparing handover orientations and quantifying object affordances. This research shows that depending on the object, natural handover orientations may not be receiver-centered; thus, robots may need to examine the quality of natural handovers before using them as a template for selecting appropriate object orientations.

The remainder of this chapter is an adaptation of a manuscript that I co-authored, entitled, “Characterization of Handover Orientations used by Humans for Efficient Robot to Human Handovers” [29]. This manuscript is published in the proceedings of the 2015 IEEE/Robotics Society of Japan (RSJ) International Conference on Intelligent Robots and Systems (IROS) held from September 28 to October 2, 2015 in Hamburg, Germany. The version of the manuscript that appears in this chapter contains several minor modifications compared to the submitted conference paper. These modifications are listed below.

- Figures have been replaced with high-resolution reproductions for enhanced clarity.
- The introduction contained in Section 4.1 has been abridged.
- Experimental conditions have been renamed for improved comprehensibility.
- Description of the experimental design in Section 4.4 has been expanded to include additional details.
- Conclusions Section 4.8 have been expanded to address how the work fits into the themes of the thesis.

Supplementary materials for this work can be found in Appendix B.
4.1 Introduction

Consideration of an object’s affordances and how it is handed over between giver and receiver has previously been revealed to significantly impact the handover interaction experience between people and robots. This work examines the handover orientation used by people - that is, how the giver of a handover interaction orients the object when handing it over. Studies have shown that by using an appropriate handover orientation, a robot can more clearly convey its handover intent, reduce task completion time, and make the human receiver feel safer [5, 22]. Building on these findings of the importance of handover orientation, the work presented in this chapter focuses on methods for determining proper handover orientations. Additionally, it assesses the feasibility of training robots to use appropriate handover orientations based on observations of natural human-human handovers.

4.2 Background

Determining appropriate handover orientations is a challenging task, as the proper handover orientation of an object not only depends on the object’s physical properties, but also factors such as receiver state [120], object function, [21] and object affordances. Furthermore, the proper orientation of an object may be non-unique due to multiple possible usages of the object [17]. In some previous studies, the handover orientation is provided to the robot a priori, [e.g., 5], while in others, a few methods for computing handover orientations have been developed. Kim et al. proposed a dual arm planner that uses user-provided object-specific affordance information to compute handover orientations [64]. Cakmak et al. performed a survey and determined orientations for a cylindrical object that better convey the robot’s handover intent [22]. Later, Cakmak et al. proposed two methods for determining handover orientations [21]: one uses user-provided examples of good and bad handover orientations, and the other plans handover orientations using a human kinematic model, while ignoring object affordances.

Existing methods often rely on user-provided object-specific information to capture the object’s affordances. To avoid the need for explicit user training and allow robots to automatically determine appropriate handover orientations, Chan et al. proposed an approach based on observing natural human handovers and
learning the handover orientations used [28]. This approach assumes that natural handover orientations used by humans are indeed appropriate. However, the nature of handover orientations used in human handovers is still not very well understood in a way that they can be applied to robot handovers. Thus, this study investigates the nature of the handover orientations used by humans towards developing a relevant mapping to selection of appropriate robotic handover.

4.3 Objectives

The objectives of this study are to collect a set of handover orientations used by humans for a set of common objects in different conditions, and determine features that describe handover orientations used in natural handovers, for guiding robot-human handover design. The research questions for this work are:

1. What are the handover orientations used by people for handing over different common objects?

2. Are there observable patterns in the handover orientations used by people?

3. Do people naturally use a handover orientation that is considerate of the receiver?

It is posited that in collaborative scenarios, objects are handed over with the intention of letting the receiver use the object to complete the task at hand. Thus, a giver should use a handover orientation that is considerate of the receiver and allows them to readily take and use the object to ensure efficient collaboration.

4.4 Object Handover User Study

4.4.1 Experiment Design

A user study was conducted where participants, consisting of giver and receiver dyads, handed over various objects while the motions and orientations of both the participants and objects were tracked through a motion capture setup (described later in Section 4.4.5). This study aims to uncover what orientations people naturally use for handing over the objects, and if the orientations used would differ
depending on the giver’s focus. Thus, three conditions are tested within participant pairs:

- **Condition A**: (Natural handovers) Participants are not given any explicit instructions on how to handover the objects.

- **Condition B**: (Giver-centered handovers) Focus is placed on the giver by asking the giver to hand the object over in a way that is the easiest and most convenient to him/herself.

- **Condition C**: (Receiver-centered handovers) Focus is placed on the receiver by asking the giver to hand the object over in a way that is the most comfortable and convenient to the receiver, giving consideration to the usage of the object and function of the different parts.

### 4.4.2 Objects

Twenty different, common objects that people were expected to hand over on a daily basis were selected to be handed over during the experiment (Figure 4.2). The objects that were chosen varied in size, weight, value, affordances, fragility and perceived danger if the object were mishandled (i.e., there would be consequences if the object was dropped or handed over inappropriately) to represent the variety of objects that assistive robots may be expected to receive and handle in a home, work or industrial setting. Selecting a range of objects ensures that the set of kinematic features detected are not specific to a particular object’s specific shape, size, weight, use, value, fragility or other intrinsic features. All of the chosen objects could be single-handedly transferred from a giver to receiver, and are considered to be within the payload capabilities of most current robots used for human-robot interaction - e.g., KUKA LBR iiWA (KUKA Robotics, Augsburg, Bavaria, Germany) [70], Willow Garage PR2 (Willow Garage, California, USA), Barrett WAM (Barrett Technology, LLC, Massachusetts, USA) - to make the results of the work immediately applicable.
Figure 4.2: Common everyday objects that were used in the handover object orientation user study. Approximate locations of retro-reflective markers on the objects are shown as blue dots. Arbitrary coordinate frames assigned to the objects are also displayed.
4.4.3 Participants

Recruitment for this experiment was advertised through the laboratory web page, mailing lists (Section B.1.2), campus bulletin board postings (Section B.1.1), and through word-of-mouth. Twenty participants (9 females, 11 males), aged 19-61 years \( M=28.15, SD=11.55 \) were recruited in total. All participants provided their informed consent prior to the experiment using the form shown in Section B.2; they were notified that their participation was voluntary, and they were allowed to withdraw from the experiment at any time. Additionally, permission was obtained from all participants to record both video and motion-capture data from the experiment. Participants were rewarded for their time with a candy bar following completion of the experiment.

4.4.4 Participant Task

For this experiment, participants worked in pairs. At the start of a set of trials one participant was arbitrarily designated the object giver, and the other the object receiver in the handover task. Prior to the start of a trial, each participant stood facing the other approximately 1.5 m apart. The experimenter placed one of the twenty objects in a stable position on one of three tables to the rear and sides of the giver (Figure 4.3); both the placement and orientation of the object was randomized. At the sound of an experimenter-given verbal ‘go’ signal, the giver picked up the object and handed it to the receiver. For consistency, all participants were asked to use their right hand, regardless of their handedness. Following the handover, the experimenter retrieved the object from the receiver and asked the participants to return to their starting positions. The experimenter, again, placed another object on one of the three tables and the task was repeated. After performing handovers of all twenty objects, the participants were given an opportunity to take a break before new instructions for the following set of trials were given. The participants then swapped giver/receiver roles and performed another set of twenty handovers. On average, the time taken for a pair of participants to complete the tasks was 1.5 hours.

To measure natural handovers without carryover effects for each pair, the first two sets of twenty handovers were condition A. The remaining four sets varied be-
Figure 4.3: Experimental setup for examining object orientations during handovers. Motion capture equipment and garments are not depicted.
tween conditions B and C in counter-balanced order. There were short pauses after each set of twenty handovers for explanation of the next condition, and optional breaks were given to prevent fatigue; however, none required breaks. A total of six sets of twenty handovers were performed per pair of participants, with giver and receiver switching roles after every 20 handovers to cover all three conditions. Thus, 1200 handovers were captured for the data set used in this study.

4.4.5 Motion-Capture System

Data for this experiment was collected using a Vicon motion-capture system MX T-160, (Vicon Motion Systems, Oxford, UK) with eight cameras. Infrared retroreflective markers were placed on the objects (Figure 4.2) and the participants (Figure 4.4) for tracking of key features with respect to an arbitrary global coordinate frame. Time-series kinematic data of these markers were captured at a rate of 300 Hz. The Vicon Nexus software (Vicon Motion Systems, Oxford, UK) was used for computing object orientation and joint angles in post processing.
Figure 4.4: Diagram of retroreflective marker placement and labels on participants. Both markers that were elected for use in this study and those that were not used are shown in different colours.
4.4.6 Hypotheses

As most objects have specific affordances (for cutting, pressing, grasping, etc.), it is plausible that if givers present objects in a receiver-centered fashion, they would orient the object to present the grasping part to the receiver; thus, there would be observable patterns in the handover orientations used in condition C. If givers present objects in a giver-centered fashion, they may simply use any arbitrary orientation that is convenient at the moment, or they may grasp an affordance that is convenient to them, such as the handle. Thus, it is expected that differences will arise between the handover orientations used in conditions B and C. Furthermore, in natural handovers (condition C), it is predicted that the giver will either use a handover orientation that is convenient to himself/herself or one that is convenient to the receiver. Stated formally, the hypotheses are as follows:

H1 There are observable patterns in the handover orientations used among participants in receiver-centered handovers (condition C).

H2 Handover orientations used in giver-centered handovers (condition B) and receiver-centered handovers (condition C) are significantly different.

H3 Handover orientations used in natural handovers (condition A) are either similar to those used in giver-centered (condition B) or receiver-centered (condition C) handovers.

4.5 Data Analyses

4.5.1 Handover Orientation Extraction

To extract the handover orientation, the hand trajectories from the motion capture data were examined. Figure 4.5 shows the typical trajectory of the giver’s and receiver’s hands in a handover trial. In the bottom plot of this figure depicting the distance between the giver’s and receiver’s hands, a distinct trough is observed which is assumed to represent the instant that the object is transferred from giver to receiver. The instant of object transfer was found by determining when this
distance between hands is minimized and the object’s orientation at this time was defined as the handover orientation.

![Charts representing typical hand trajectories of giver and receiver](image)

**Figure 4.5:** Charts representing typical hand trajectories of giver and receiver in Cartesian coordinates. The instant of object transfer is found by locating the minimum distance between giver’s and receiver’s hands.

An arbitrary coordinate frame was assigned for each object (shown in Figure 4.2). To be able to translate the results to different locations in space, the handover orientations were computed in a base frame defined by the giver and receiver locations. The ground surface normal is treated as the base frame z-axis and the vector pointing from the receiver’s to the giver’s torso was chosen to be the x-axis; y-axis assignment was automatically completed by enforcing a right-handed coordinate frame (Figure 4.6).

After extracting the handover orientations from all participants, the mean handover orientation $\bar{R}$ was calculated for each object using Equation 4.1:

$$\bar{R} = \arg\min_R \sum_i d(R, R_i)$$  \hspace{1cm} (4.1)

where $R_i$ is the handover orientation extracted from participant $i$ for the object, and $d(R, R_i)$ to be the positive value of the angle of rotation between $R$ and $R_i$ as mathematically described in Equation 4.2;
Figure 4.6: Base coordinate frame used during examination of object orientations during handovers as defined by giver and receiver locations.

\[ d(R, R_i) = \arccos \left( \frac{\text{trace}(R^{-1}R_i) - 1}{2} \right) \]  \hspace{1cm} (4.2)

where \( \text{trace}(R^{-1}R_i) \) is the trace of the matrix \( R^{-1}R_i \). Equation 4.2 essentially converts the rotation matrix \( R^{-1}R_i \) into the angle component of the axis-angle representation of the rotation. Thus, Equation 4.1 finds a mean rotation \( \bar{R} \) that minimizes the sum of rotation angles between \( \bar{R} \) and each measured \( R_i \). The MATLAB (MathWorks, Natick, Massachusetts, USA) built-in \texttt{fminsearch} utility was used to optimize Equation 4.1.
4.5.2 Affordance Axes

From preliminary inspection of the data, participants tended to display patterns in handover orientation for most objects. For example, when handing over the mug, participants tended to keep the mug upright, and when handing over the hammer in condition C, most participants orientated the handle towards the receiver. It appeared that participants tended to align a specific axis associated with the object in the same general direction when handing it over. To help classify this phenomenon, the notion of an affordance axis was created. It is hypothesized that for most objects, there is an affordance axis $\phi_{Aff}$ associated to the object, such that when people hand over the object while giving consideration to the object’s affordances, they orient the object to align $\phi_{Aff}$ in a certain direction. Intuitively, this would be an axis along the handle for the hammer, and an axis oriented perpendicular to the bottom surface for the mug. Mathematically, $\phi_{Aff}$ is defined as follows: let $\hat{\phi}_{Aff}$ be a unit vector pointing along $\phi_{Aff}$. $\hat{\phi}_{Aff}$ is computed using Equation 4.3:

$$\hat{\phi}_{Aff} = \arg\min_{\hat{\phi}} \sum_{i} \arccos(\bar{R}\hat{\phi} \cdot R_{i}\hat{\phi})$$ (4.3)

In Equation 4.3, $\bar{R}\hat{\phi}$ and $R_{i}\hat{\phi}$ rotates a unit vector $\hat{\phi}$ in the object frame by $\bar{R}$ and $R_{i}$ respectively to obtain the vector in world frame. If $\hat{\phi}$ indeed gives the affordance axis of the object, then the vectors $\bar{R}\hat{\phi}$ and $R_{i}\hat{\phi}$ should align. Thus, $\hat{\phi}_{Aff}$ is found by finding the $\hat{\phi}$ that minimizes the sum of angles between $\bar{R}\hat{\phi}$ and each $R_{i}\hat{\phi}$. The affordance axes were from the receiver-centered handover orientations extracted from condition C, and the MATLAB fminsearch function was used to perform the optimization.

4.5.3 Patterns in Handover Orientations

In Section 4.4.6, it was hypothesized that there are observable patterns in the handover orientations used in condition C (H1). To find support for this hypothesis, the angles between $\bar{R}\hat{\phi}_{Aff}$ and $R_{i}\hat{\phi}_{Aff}$ for each object, labelled $\theta_{i}$, were computed and plotted in histograms. If H1 were false, meaning that there are no observable patterns among the handover orientations, and $R_{i}$ were randomly distributed in rotation space, then $R_{i}\hat{\phi}_{Aff}$ would not be in general alignment with $\bar{R}\hat{\phi}_{Aff}$, and
the histogram of $\theta_i$ would be distributed across all angles. A Monte Carlo simulation of twenty random handover orientations indeed confirms this expectation (Figure 4.7). Kuiper's test was used to determine if the distribution of $\theta_i$ for each object is different from a uniform distribution.

![Histograms showing distribution of $\theta_i$](image)

**Figure 4.7**: Four repetitions of Monte Carlo simulation results showing histograms of $\theta_i$, the angle between $\hat{\phi}_{Aff}$ and $\hat{\phi}_{Aff}$. 20 random handover orientations were generated for each simulation repetition. A spread of $\theta_i$ among all angles can be seen.

### 4.5.4 Comparison of Handover Orientations Across Conditions

To compare handover orientations used across conditions and test hypotheses H2 and H3, differences between handover orientations used were examined for each condition. Let $R_U^i$ and $R_V^i$ be the handover orientations used by participant $i$ for conditions $U$ and $V$ respectively. The measurement of difference, $\theta^{UV}_i$, between these two orientations is computed as the angle between the affordance axes in these orientations as described in Equation 4.4:

$$
\theta^{UV}_i = \arccos(R_U^i \hat{\phi}_{Aff} \cdot R_V^i \hat{\phi}_{Aff})
$$

If the handover orientations used in conditions $U$ and $V$ are similar, then the mean angular difference, $\bar{\theta}^{UV}$, should be small and less than some $\delta$. Note that $\delta$ is not equal to zero, due to natural variations in human motion. To estimate $\delta$, the average spread in the angle between the affordance axes measured in condition $C$ was computed using Equation 4.5:
\[ \delta = \sum_{i} \arccos(\hat{R}_C^C \hat{\phi}_{\text{Aff}} \cdot \hat{R}_i^C \hat{\phi}_{\text{Aff}}) \]

(4.5)

where \( n \) is the total number of participants. To determine if the handover orientations between conditions were different, t-tests were performed to determine if \( \bar{\theta}^{UV} \) is different from \( \delta \). Significant results are reported at an \( \alpha \) level of 0.05 with Bonferroni correction.

### 4.6 Results

#### 4.6.1 Handover Orientation and Affordance Axis

To examine the data, the extracted handover orientations were plotted in 3D plots. The left chart in Figure 4.8 shows an example from a teapot handover trial in condition C. Referring to the teapot’s assigned coordinate frame shown in Figure 4.2, in this trial, the teapot is handed over upright, with the handle pointed towards the receiver, slightly to their right. The right chart in Figure 4.8 shows the handover orientation from all trials. The computed mean orientation \( \bar{R} \) is shown as the bold, dotted coordinate frame, and the long thin line shows the computed affordance axis \( \hat{\theta}_{\text{Aff}} \) in the mean orientation frame. For visualization and comparison across conditions, Figure 4.9 shows the handover orientations from all trials, the mean orientation, and affordance axis, for all objects in all conditions.
Figure 4.8: Figure showing orientation of a teapot for one and multiple trials as it is being handed over from giver to receiver. The left image shows an example teapot handover orientation from condition C for a single trial. Red, green, and blue lines represent X, Y, and Z axes respectively. In the right image, teapot handover orientations for all trials from condition C. The thick, dotted red, green, and blue lines denotes the computed mean $\bar{R}$, and the long thin line represents the computed affordance axis $\hat{\phi}_{Aff}$ in mean handover orientation frame $\bar{R}$. 
**Figure 4.9**: Handover orientations employed during human-human handover for all objects in all conditions. Bold, dotted coordinate frames represent mean orientations and long thin lines represent computed affordance axes.
4.6.2 Patterns in Handover Orientations

For a few objects, it is quite easy to see from Figure 4.9 that there is a clear pattern in the handover orientations. For example, for the mug and teapot, it can be observed that most trials have a handover orientation where the z-axis of these objects are closely aligned with each other. Indeed the computed affordance axes of these objects lie roughly along the z-axis. To facilitate the recognition of such patterns in other objects as well, histograms were plotted of $\theta_i$ for each object in condition $C$ in Figure 4.10. Kuiper’s test shows that for all objects, the distribution of $\theta_i$ is significantly different from a uniform distribution (knife: $p = 0.008$, pen: $p = 0.009$, all other objects: $p < 0.005$).
Figure 4.10: Histograms of the angles between $\vec{R}_{\hat{\Phi}}^{\text{Aff}}$ and $R_i\hat{\Phi}_{\text{Aff}}$ (i.e., $\theta_i$) for each object in condition $C$ during the study examining handover orientations.
4.6.3 Comparison of Handover Orientations Across Conditions

Table 4.1 shows the computed means and standard deviations for $\bar{\theta}_{AB}$, $\bar{\theta}_{AC}$, and $\bar{\theta}_{BC}$, as well as the t-test results, with statistically significant results highlighted. The results varied by object. For the book, mug, and plate, comparisons did not yield significant differences in handover orientations between all conditions. For the bottle, camera, fork, hammer, pen, remote, scissors, screwdriver, wineglass, and wrench, comparison across all conditions were significant suggesting that a different handover orientation was used for each condition. For the remainder of the objects, comparisons between condition pairs yielded mixed results.
Table 4.1: Results of comparison of handover orientations between conditions. Mean difference ($M$), standard deviation ($SD$), $p$-values ($p$), and $t$-statistics ($t(19)$) are presented. Significant results from t-tests are highlighted.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Book</th>
<th>Bottle</th>
<th>Camera</th>
<th>Cereal Box</th>
<th>Umbrella</th>
<th>Flowers</th>
<th>Fork</th>
<th>Hammer</th>
<th>Knife</th>
<th>Mag</th>
<th>Pen</th>
<th>Plate</th>
<th>Remote</th>
<th>Scissors</th>
<th>Screwdriver</th>
<th>Stapler</th>
<th>Teapot</th>
<th>Tomato</th>
<th>Wineglass</th>
<th>Wrench</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>22.7</td>
<td>18.3</td>
<td>28.3</td>
<td>35.7</td>
<td>42.4</td>
<td>26.0</td>
<td>27.9</td>
<td>36.7</td>
<td>42.3</td>
<td>11.1</td>
<td>50.1</td>
<td>6.5</td>
<td>18.6</td>
<td>29.8</td>
<td>38.3</td>
<td>47.5</td>
<td>7.9</td>
<td>13.5</td>
<td>11.3</td>
<td>27.4</td>
</tr>
</tbody>
</table>

| $\bar{\theta}$ | $M$ | 34.4 | 32.6 | 53.7 | 65.8 | 81.9 | 39.5 | 52.7 | 99.1 | 72.2 | 19.9 | 82  | 13.8 | 31.6 | 74.4 | 78  | 68.8 | 11.3 | 31.2 | 33.8 | 88.4 |
| $SD$ | 41.8 | 24.5 | 39.5 | 42.5 | 53.4 | 28.3 | 39.2 | 58.5 | 45.1 | 28.3 | 51.2 | 23  | 18.2 | 42.3 | 60.9 | 53.0 | 8.7  | 22.7 | 34.8 | 44.9 |
| $p$ | .112 | .008 | .005 | .003 | .002 | .023 | .005 | .000 | .004 | .009 | .006 | .084 | .002 | .000 | .004 | .045 | .047 | .001 | .005 | .000 |
| $t(19)$ | 1.25 | 2.62 | 2.87 | 3.17 | 3.31 | 2.13 | 2.83 | 4.76 | 2.96 | 1.39 | 2.79 | 1.43 | 3.21 | 4.72 | 2.91 | 1.79 | 1.76 | 3.48 | 2.88 | 6.08 |

| $\bar{\theta}$ | $M$ | 28.1 | 30.8 | 46.9 | 40.4 | 71.7 | 42.3 | 47.5 | 67.8 | 54.4 | 12.6 | 94.8 | 8.4  | 40.8 | 56.0 | 58.6 | 96.7 | 10.8 | 20.3 | 20.9 | 65.9 |
| $SD$ | 41.0 | 21.2 | 26.4 | 26.6 | 52.7 | 22.5 | 25.7 | 57.2 | 50.3 | 8.2  | 50.8 | 5.5  | 39.7 | 42.0 | 37.8 | 56.4 | 7.2  | 14.0 | 16.9 | 54.2 |
| $p$ | .283 | .008 | .003 | .217 | .011 | .002 | .001 | .013 | .148 | .218 | .000 | .071 | .011 | .006 | .013 | .000 | .043 | .021 | .010 | .002 |
| $t(19)$ | 0.59 | 2.62 | 3.15 | 0.8  | 2.49 | 3.23 | 3.41 | 2.43 | 1.07 | 0.79 | 3.93 | 1.53 | 2.51 | 2.79 | 2.4  | 3.9  | 1.81 | 2.18 | 2.52 | 3.17 |

| $\bar{\theta}$ | $M$ | 46.4 | 35.8 | 46.3 | 62.2 | 65.2 | 53.0 | 50.8 | 95.2 | 88.3 | 19.2 | 90.8 | 12.5 | 35.1 | 74.8 | 79  | 71.8 | 10.7 | 26.3 | 32.6 | 97.8 |
| $SD$ | 53.1 | 28.6 | 33.8 | 39.5 | 46.1 | 31.8 | 39.8 | 58.1 | 61.8 | 23.8 | 55.9 | 21.0 | 32.1 | 44.9 | 58.2 | 59.6 | 3.9  | 20.6 | 31.6 | 59.3 |
| $p$ | .030 | .007 | .014 | .004 | .020 | .001 | .009 | .000 | .002 | .072 | .002 | .107 | .016 | .000 | .003 | .042 | .002 | .006 | .004 | .000 |
| $t(19)$ | 2.0  | 2.73 | 2.38 | 3.0  | 2.22 | 3.79 | 2.57 | 4.5  | 3.33 | 1.52 | 3.25 | 1.29 | 2.31 | 4.48 | 3.13 | 1.82 | 3.2  | 2.78 | 3.02 | 5.32 |
4.7 Discussion

4.7.1 Patterns in Handover Orientations

Comparing Figure 4.7 and Figure 4.10, it can be observed that for most objects, the distribution resulting from condition C clearly differs from those generated by random orientations given in Section 4.5.3. In particular, the book, bottle, camera, mug, plate, remote, teapot, tomato, and wineglass exhibit a pronounced peak near 0°, with equal to or more than 50% count falling in the smallest bin of $\theta_i < 10^\circ$. For the rest of the objects, except the pen, although less pronounced, a skewed distribution towards 0° can still be seen. This suggests that participants do have a preference on aligning the affordance axis of the objects in a certain direction, and thus support is found for the claim in H1. Although as a first step, $\theta_i$ has been tested using the Kuiper’s test to compare with a simple uniform distribution and obtained significant results for all objects, additional investigation is required to determine whether a uniform distribution indeed represent the distribution of $\theta_i$ in the random case. Furthermore, rotational symmetry in some of the objects may also affect how the random case distribution looks like. Further work is needed to draw stronger conclusions.

4.7.2 Comparison of Handover Orientations Across Conditions

T-test results in Table 4.1 for $\bar{\theta}^{BC}$ comparing handover orientations between conditions B and C show significance for a majority of the objects, including the bottle, camera, cereal box, flowers, fork, hammer, knife, remote, scissors, screwdriver, teapot, tomato, wineglass, and wrench. This suggests that for these objects, the givers use a different handover orientation depending on whether they are focusing on their own comfort (condition B) or on the receiver’s comfort (condition C). Indeed, with a closer inspection of the data and experiment videos, there are observable difference in the handover orientations used. For example, with the bottle, participants orient it upright, but with a wide variation in the tilt direction in condition B. In condition C, there is less variation and participants tend to tilt the bottom towards the receiver. Similarly, with the hammer and the wrench, most participants oriented the hammer head and wrench head towards the receiver in
condition B, but presented the handle in condition C. At this point, it is also worth noting that the computed mean orientations captures such characteristics quite well for these objects as well as others. Thus, it is possible to teach the handover orientations extracted from condition C to robot givers, allowing them to perform receiver-centered handovers. Based on these results, support has been for hypothesis H2 for the above objects. The handover orientation used by the giver changes depending on their focus during the handover.

Although the statistical tests for the mug were not significant, it was observed that in more than 50% of the trials within condition B, the giver either grabs onto the handle or the rim without presenting the handle to the receiver; whereas in condition C, the handle was presented to the receiver in over 75% of the trials. This difference in the handover configuration between conditions B and C is indeed captured by the computed mean orientations. The t-test result indicates that the affordance axis of the mug (computed to be along the z-axis) aligns in the same direction, meaning that even though the givers were asked to focus on their own comfort in condition B, they still tended to keep the mug upright. This is likely due to the associated potential consequence of spilling, if the cup were filled with liquid.

Comparing the handover orientations used in condition A with those in conditions B and C (Table 4.1 \( \bar{\theta}^{AB} \) and \( \bar{\theta}^{AC} \)), results revealed no significant difference between the handover orientations used between conditions A and B, and between A and C for the book, mug, plate, and teapot. Thus, support for H3 was found, showing that for these objects the handover orientation in condition A is similar to those in conditions B and C. The mug, plate, and teapot all have the associated risk of spilling if they had contents. Hence, in all conditions, \( \geq 85\% \) of the time the giver kept the object upright. As for the book, people normally read a book from the beginning. Thus, the giver seems to naturally orient the front cover upwards in all conditions.

For the cereal box, knife, and tomato, t-test results show that handover orientations used in condition A were similar to that of condition C, but different from condition B. When handing over a knife, there is a possibility of injuring the receiver if an inappropriate orientation is used. Thus, in condition C givers tend to orient the knife tip away from the receiver, and naturally, in condition A, they use a
similar orientation. In condition $B$, however, when the givers were explicitly asked to focus on their own comfort, they seem to then change the way they orient the knife. A change in handover orientation for the mug and plate was not seen - it is predicted that this is due to the giver knowing that if the mug or plate were not held upright, any contents would be sure to spill. However, in the case of handing over the knife, as long as the receiver exercises extra caution, injury can still be avoided.

The results of this study shows that for the bottle, camera, umbrella, fork, hammer, remote, scissors, screwdriver, wineglass, and wrench, handover orientations in condition $A$ were different from those in conditions $B$ and $C$. For some of these objects, namely the hammer and wrench, it appears that this is because the orientations used in conditions $A$ and $B$ are quite widely distributed. This perhaps suggests that these objects do not have a strong enough affordance characteristic that would prompt givers to hand them over in any particular orientation unless asked explicitly to consider the object’s function. For some other objects such as the remote, it appears that in condition $A$, there is a mix of handover orientations from conditions $B$ and $C$, suggesting that depending on the individual, some givers naturally use a handover orientation that is giver-centered, and some, receiver-centered.

4.7.3 Implications Towards Building Intelligent Robots

The end goal of this work is to allow robots to learn receiver-centered handover orientations from observing natural handovers. The work presented in this chapter has contributed towards this goal by introducing the idea of the object affordance axis, and providing a method for computing mean handover orientations. The study presented here has shown that handover orientations used in natural handovers may not all be receiver-centered. Thus robots may need a method for distinguishing them. In the cases where natural handover orientations differ from receiver-centered orientations, there seems to be a wide variation across the handover orientations used. Therefore, by identifying the affordance axis and computing the mean or variance of $\theta_i$, the robot may be able to determine whether the observed natural handover orientations are receiver-centered.

This work provides an optimization based method for calculating mean handover orientations. This method can be used by robots to compute the proper han-
dover orientation from a set of observed handover orientations. Using this method, computation of handover orientations for a set of common objects in different conditions was performed. The mean orientations calculated from the receiver-centered handovers could potentially be used by robots when handing over the objects to people to facilitate more efficient and socially acceptable interaction.

4.8 Conclusions

The investigations presented within this chapter provides building blocks towards allowing robots to interpret human nonverbal cues to infer object attributes, i.e., part of the second research question of this thesis. Specifically, the aim is to have robots be able to orient handover objects appropriately (in terms of the objects’ affordances) to human receivers. This work adopts a scheme of having robots trained on the appropriateness of handover orientations via observations of human-to-human handovers. To this end, a study was conducted with the intention of determining the suitability of using human-to-human handover examples to train robot givers on how to appropriately orient handover objects based on the objects’ affordances, while simultaneously developing the necessary tools to enable robots to learn how to appropriately orient objects during robot-to-human handovers using these observations.

Towards enabling robots to learn proper handover orientations from observations, this work introduces the idea of object affordance axes, and an optimization based method from computing mean handover orientations. Through a user study, handover orientations used by humans were surveyed for a set of objects in different conditions. Mean orientations were computed from the data, and extracted mean orientations from receiver-centered handovers could potentially be used to teach robots how to appropriately hand over these objects to people to promote more effective cooperation. Through comparison of three structured handover conditions, natural handover orientations were found to differ from receiver-centered orientations for some objects. Thus when a robot tries to learn from natural handovers, it may need to consider the quality of the observed handover orientations. Measurement of the variation in the observed handover orientations may provide some kind of a quality measurement, as objects that have different natural and
receiver-centered handover orientations appear to exhibit larger variance in the natural handover orientations.

In relation to the second research question of this thesis, the results of this work suggest that the aim of having robots observe, learn, and/or mimic behaviours from human nonverbal cues may be confounded by variation in these observations. Factors such as personal preferences, environmental configuration and cultural aspects will, with no doubt, impact how humans carry out handovers with respect to the nonverbal cues that are used. Thus, having a robot learn by example may require a constrained or artificial data set - e.g., only learning handover orientations using the receiver-centered handover data set. Alternatively, as previously mentioned, the quality of the dataset may need to be assessed to sort good examples from sub-par ones. Future work includes running user studies to confirm that the extracted receiver-centered mean orientations indeed improve robot-human handovers and facilitate better interaction, as well as devising methods for identifying receiver-centered handover orientations from natural handovers.

Continuing the theme of recognizing and interpreting nonverbal cues conveyed by a human, the following chapter examines how a robot receiver can learn to detect a human giver’s intent to handover an object. In that work, a machine learning approach is used to both automatically identify important kinematic cues indicating handover intent as well as detect when a human giver intends to handover an object to a receiver. The work serves as a first step or building block in allowing a robot to receive an object from a human counterpart.
Chapter 5

Automated Detection of Handovers using Kinematic Features

Previously, Chapters 3 and 4 have examined how nonverbal cues and proper handover object orientation may improve aspects of a robot-to-human handover. These works have examined and discussed how a robot giver could leverage object affordances and gaze cues when handing over an object to a human receiver to make the experience more fluent and efficient. The rest of the works contained within this thesis will study the role of nonverbal cues in handover interactions where the roles of the human and robot are opposite to that examined in the previous two chapters; that is, where the human acts as the giver and the robot as receiver (human-to-robot handover).

In establishing what prior work has been done in the area of human-to-robot handovers, there appears to be little on the subject; rather, much of the literature on handovers between human and robot agents explore robot-to-human handovers. Thus, having somewhat of a clean slate to work with, the research described in the next few chapters attempts to establish the basics of the interaction while also probing the available design space.

This chapter continues the thread picked up in Chapter 4 and focuses on having robots recognize and interpret nonverbal cues generated by a human during
handovers - i.e., the second research question posed in this thesis. Here, one of the first steps in what a robot needs to do when receiving an object from a person is examined, namely, recognizing the intent of that person to handover an object. To achieve automated detection of handovers, a machine learning approach is utilized for recognizing handover intention from kinematic motions of the human giver. Through a structured procedure of selecting 22 kinematic features based on predictive value for handover intent, machine learning models were able to be developed which could detect handover motions with 97.5% accuracy. Given that a motion capture system was used for collecting kinematic data from human-human dyads - a technique that would most likely not be available for robots in the field - a pilot study conducted using a more typical sensor used in unstructured environments (the Microsoft Kinect version 2) was also performed. The results are encouraging, demonstrating considerable potential and feasibility of this method for detecting handover intent (and perhaps other gestures for human-robot interaction) using kinematic features.

This work presented here has been published in The International Journal of Robotics Research (IJRR) volume 36, issue 5-7, a special issue on HRI in 2017. I co-authored the original manuscript, entitled, “Automated Detection of Handovers using Kinematic Features” [96]. The data and results which appeared in this work were derived from the same data used in the previous chapter (Chapter 4). Several changes to this manuscript were made for it to be included in this thesis which include:

- Figures have been replaced with high-resolution reproductions for enhanced clarity.
- The introduction of the manuscript (Section 5.1) has been abridged.
- Background information that has already presented in Chapter 2 has been removed.
- Details of the experimental setup of this work is the same as that which appears in the previous chapter (see Section 4.4) and has been removed.

Supplementary materials for this work can be found in Appendix B.
5.1 Introduction

While seemingly effortless for most humans, recognizing a handover motion and receiving the object in an appropriate manner from a giver is a challenging task for a robot assistant. A key skill enabler is the ability to detect nonverbal cues from the giver in order to infer the timing and location of the handover.

This chapter presents work which was done to develop building blocks towards enabling robots to robustly and naturally receive an object from a human giver: namely, to detect when a human giver initiates a handover of an object through recognition of nonverbal kinematic cues. This work is predicated on the recent availability of inexpensive, off-the-shelf sensors such as the Microsoft Kinect (Microsoft, Redmond, Washington, USA), providing ready access to human body kinematic data from which cues relating to handovers and other interactions can be detected. In essence, the aim of this work is to examine the feasibility of using data from such devices for this purpose through an exploratory study.

This work is a departure from prior work in that it attempts to recognize the giver’s intent to hand over objects through machine learning of kinematic data obtained from the giver rather than performing recognition on image or manually coded data. The advantage of this method is that it allows online handover detection using off-the-shelf depth sensors that are able to perform skeleton tracking. With this method of detection, a robot would be able to differentiate between handovers aimed towards it as the intended receiver and those directed at other potential receivers. Furthermore, a robot would not need to rely on verbal cues or the observation of an object in the giver’s hand to detect that a handover is taking place.

In the following section, prior research is reviewed in the areas of nonverbal cues, proxemics, kinematics, and the use of machine learning in recognizing handovers (Section 5.2). This is followed by a description of the objectives and approach of this study (Section 5.3), experimental methods used to collect the data (Section 5.4) and how the data was analyzed (Section 5.5). An overview of the results is then presented (Section 5.6) followed by a discussion of the impact of these results (Section 5.7). The chapter is then concluded by discussing the implications, limitations, and areas for future examination of this work (Section 5.8).
5.2 Prior Work

The work presented in the paper is derived from several fields of research: non-verbal cues, proxemics, and kinematics in handover tasks, and machine learning. An overview of prior work in these areas is presented here.

5.2.1 Non-Verbal Cues, Proxemics and Kinematics in Handovers

Prior works have demonstrated that nonverbal cues play a significant role in coordinating handovers between human participants [43]. Recent studies investigating nonverbal gestures involved in handovers have examined more subtle cues such as gaze and eye contact. This work demonstrated that where participants look during handover plays a significant part in the coordination of handovers [83, 88, 112, 113, 123]. Work on social gaze for robots by Mutlu et al. provides some guidance on how gaze can be used for such coordination [92]. In addition to gaze, grip and load forces have been shown by Chan et al. and Kim and Inooka to play a significant role in the coordination of handovers, and have led to the observation that the giver seems to be responsible for the safety and successful handover of the object, while the receiver is responsible for the timing of the handover [26, 63].

Much of the early work in robot-to-human handovers relied on proxemics studies by Hall, whose work suggested that distances between persons interacting socially are influenced by culture, attitudes, social standing, and relationships to one another [45]. A later study showed that interpersonal distance between human givers and receivers at the moment an object is handed over varied considerably, with the average distance being 1.16 m away - roughly the distance where both giver and receiver must have arms outstretched during handover [13]. To compare this to proxemics between robots and humans, Walters et al. found that most of their participants allowed a robot to approach to within a personal distance (0.6 to 1.25 m), suggesting that humans may treat robots similar to humans socially [121]. However, 40% of their participants allowed a robot to approach to within an intimate distance of 0.5 m, implying that humans are more tolerant of robotic close encounters, which would be perceived as over-familiar or threatening in a similar human-human context.
In order to better understand the spatial behaviours of participants within human-human handovers, several studies have considered where handovers occur in the spatial domain \([13, 54]\), and the joint/limb dynamics of both the giver and receiver during handover \([41, 60, 106]\). In Sisbot and Alami’s work, the authors use kinematic features, along with preferences and gaze of the human receiver, to help a robot giver plan trajectories to navigate to a handover location safely and in a socially comfortable manner \([106]\). Kajikawa et al. have determined that handovers between humans share several common kinematic characteristics such as a rapid increase in the giver’s arm velocity at the start of the handover \([60]\). They also observed the occurrence of a delay in handover reaches, as the receiver begins their reach for the object only after the giver achieves a maximum approach velocity. Basili et al. quantified this delay by observing that nonverbal indicators that a handover was to occur happened approximately 1.2 s prior to the occurrence of the actual handover \([13]\).

These observations are particularly important for the development of a handover controller as robots may be able to anticipate rather than react to handovers, leading to more fluent interactions \([50]\). That is, if nonverbal cues indicating an object is being passed can be observed early enough, a robot may be able to act quickly enough to allow the handover to proceed seamlessly and fluently from the perspective of a human giver.

### 5.2.2 Machine Learning for Handovers

The use of machine learning for classifying kinematic data has been a recent area of interest. Machine-learning approaches have been successful in recognizing gait patterns \([14, 46, 73]\), physical gestures \([34, 51, 52]\) and other day-to-day physical activities \([80, 93]\). For example, Takano et al. presented a hidden Markov model based method for learning and recognizing interactive behaviours from motion-capture data used to generate interactive robot behaviour \([114]\). More recently, Gaussian mixture models have become more common in human gesture recognition tasks \([99, 109, 110]\), achieving varying degrees of classification accuracy from 70% up to 90%.
A few studies have utilized database searching and machine learning to assist in object handovers between robots and humans. Yamane et al. developed a database of human-human handover motions used to synthesize robot receiving motions [122]. Online performance of searching depended on the number of features being tracked and the length of observation window. In a parallel vein of work, Strabala et al. explored the use of machine-learning classifiers in recognizing the intent of a giver to hand over an object [112]. In their work, sequenced features such as orientation of the giver with respect to the receiver, eye gaze location, giver hand occupancy (holding or not holding an object), handover signals (e.g., extension of the arm), and inter-subject distances were used to recognize the intent of a human giver to hand over an object to a human receiver. These features were manually coded from video recordings.

The approach taken in this work differs from prior work in several ways. In this work, machine learning is applied to automatically select features taken from each frame of a set of time-series kinematic data. Those features that are expected to offer the most utility in detecting giver-initiated handovers form the input space for a machine learning classifier to make independent predictions at each time step. In contrast with how handover motions are recognized by Yamane et al. in [122], no search of a database or pattern recognition over a sequence of frames is required by the method to perform the detection. Additionally, unlike the work of Strabala et al. [112], the approach of this work is agnostic to whether or not the giver’s hand is occupied with an object. This is greatly beneficial as object sensing and identification is a challenging task: unless the object is known to the vision system and is well-behaved (i.e., does not deform or change properties such as colour and reflectance), vision systems typically have difficulty in detecting an object, especially if it is partially or fully occluded by the giver’s hand. In Chapter 4, such issues with object identification and tracking were avoided as a motion capture system was used: objects could be robustly and uniquely tracked as long as three or more retroreflective markers of a constellation for each object were visible to an array of cameras. However, using motion capture systems to track handover objects in the field is not feasible, particularly if there are a large number of objects, since every object to be tracked will need to be fitted with a unique pattern of intrusive markers.
The approach proposed by this work avoids these issues by observing the giver’s behaviours rather than the handover object. However, an obvious limitation of this approach is that the classifier may interpret non-handover gestures as handovers (e.g., handshakes and fist-bumps). There are multiple strategies to mitigate this limitation by coupling the system with other sub-systems which can recognize and recall an object pick-up gesture by a person, or use a vision system to detect the appearance/disappearance of objects to calculate probabilities that a potential giver picked up an object to be handed over. Investigation of these systems, however, are beyond the scope of this work.

5.3 Objectives and Approach

The objectives of this study were to collect a set of sample kinematic data used by givers during the handover of various common everyday objects, and to determine a set of features that can be used to detect the occurrence/initiation of a handover. The research questions of this work are:

1. What is the set of kinematic features of givers that could be used to detect the initiation of a handover (from the receiver’s perspective)?

2. How well can these features be used by a machine-learning approach to detect the occurrence of a handover? In particular, how does the performance of this approach compare with approaches used in prior work [e.g.,[112]]?

In this context, accuracy, sensitivity, specificity, and precision as defined as follows: accuracy is the proportion of correct binary handover/non-handover motion classifications among all observations examined. Sensitivity (also known as true positive rate) is the ability of the classifier to correctly classify data indicative of handover intent. Specificity (also known as true negative rate) is the ability of the classifier to correctly classify data indicating lack of handover intent (non-handover). Finally, precision (also known as positive predictive value) is the proportion of all correctly classified data indicative of handover intent among correct and incorrectly classified data indicating handover intent.

To achieve learning of handover cues from kinematic data, the use of SVMs is proposed. SVMs are recognized as a robust method in pattern recognition and
classification; they have been applied to numerous classification and regression
problems with exceptionally good performance [15]. There are several advantages
to using SVMs for classification tasks including the exploitation of the kernel trick
(being able to use non-linear functions for classification efficiently) and inclusion
of regularization and a large margin for better generalization. Their robustness in
binary classification and previous success in other work classifying kinematic data
- gait in particular [14, 39, 73] - makes SVMs a good fit for this work. Hence, SVMs
is applied in the study reported here for automated recognition of when a handover
occurs using human body kinematics.

5.4 Experimental Setup

The data used for this study was collected within the same experiment pre-
sented in Chapter 4 on the orientations of objects during handovers. Thus, expla-
nations and details regarding the experimental setup can be found in Section 4.4
and will not be repeated here. For this study, only motion data of the markers lo-
cated on the giver in each trial (see Figure 4.4) was analyzed; motions of the object
and receiver were not examined.

5.5 Data Analysis

5.5.1 Data Post-Processing

Post-processing of the motion-capture data for participants’ motions during
the experiment was done using Vicon Nexus software version 1.8.5 (Vicon Motion
Systems, Oxford, UK) and MATLAB (MathWorks, Natick, Massachusetts, USA).
Markers were automatically identified within the data by performing model match-
ing against template skeleton and object marker models. Gaps in marker data where
one or more markers temporarily disappeared from the view of the cameras (usu-
ally caused by occlusion of the marker) were semi-manually filled using splines or
kinematic information from other markers.

As a large amount of data was obtained (~2 million observations), the kine-
matic data recordings were downsampled from 300 Hz to 30 Hz to allow for faster
data processing and manipulation. This reduction in sampling rate also conve-
niently allows the data capture rate to be roughly the same as with off-the-shelf
motion trackers such as the PrimeSense 3D Sensor (PrimeSense, TelAviv, Israel); Intel RealSense (Intel, Santa Clara, California, USA); Kinect Version 2 (Microsoft, Redmond, Washington, USA). To reduce noise in marker trajectories, the data was
passed through a fourth-order low-pass Butterworth filter with a 6 Hz cut-off fre-
quency (the approximate frequency limit of human limb motion).

To limit redundancy in the data, a number of markers were ignored. These in-
cluded markers on the head except for the forehead (FHead) marker, any marker
located mid-segment between two joints (e.g., markers located on the collarbone,
upper arm and forearm), the Back marker that would be occluded from the perspec-
tive of the receiver during a handover, and those markers that were felt to be fairly
invariant with respect to other included markers (e.g., LChest, RChest, LWrist, and
RWrist).

Since the end goal of this work is for a robot be able to recognize that it is at the
receiving end of a handover, the data was required to represent motions from the
perspective of the receiver. This was achieved by remapping all of the data from
the arbitrary global coordinate frame of the motion capture system into an estab-
lished receiver-based frame. It was determined that the origin of this frame was
to be located halfway between the initial positions of the back and chest markers
of the receiver. The X axis vector points towards the receiver’s right, the Y axis
points forward towards the chest marker, and the Z direction points up (Figure 5.1).
It should be mentioned that although the data processing performed here is accom-
plished offline or semi-manually (e.g., the filling in of missing marker data), all of
this processing can be performed online as well (as shown in Section 5.7.3).

5.5.2 Feature Extraction

Once the giver’s kinematic data had been extracted, filtered and transformed,
MATLAB scripts were developed for the computation of several classes of fea-
tures. These features are categorized in Table 5.1. In total, 176 unique features
were generated for consideration at each time step. All features considered here
consist of filtered data from the motion-capture system, or are procedurally gener-
ated based solely upon this data (i.e., features that are not directly drawn from the
Figure 5.1: Diagram showing giver passing a book to the receiver (top) and corresponding motion-captured scene (bottom). The coordinate system seen superimposed on the receiver motion-capture model represents the one used for calculating features.
motion-capture system are generated by performing calculations upon the filtered motion-capture data). No manual coding of features is performed and all features can be generated online, significantly increasing the viability and extendibility of the method to the online detection of handovers. Thus, these features can be generated for any skeleton model containing similar markers - such as the shoulder, neck, hands, chest and head - that current sensing packages such as the Microsoft Kinect and Intel RealSense track directly out-of-the-box.

Table 5.1: Features extracted from kinematic data obtained from the motion capture system during the study investigating the automated detection of handovers.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>Euclidean and axis component distances between two markers</td>
</tr>
<tr>
<td>Receiver-based kinematics</td>
<td>Kinematics examining how the vector between a specified marker and the receiver frame origin changes (created to facilitate assignment of directionality to magnitude kinematic measures)</td>
</tr>
<tr>
<td>Linear Kinematics</td>
<td>Magnitude and axis-component kinematic measures (e.g., velocity in the Y-axis and acceleration in the Z-axis)</td>
</tr>
<tr>
<td>Angular Kinematics</td>
<td>Joint angular features considered with respect to the simple angle between two plane-defining vectors that form a joint in both simple angle and in quaternion form, and segment angular features computed between limb segments and basic coordinate planes</td>
</tr>
</tbody>
</table>

5.5.3 Labelling

To distinguish between handover and non-handover motions, the start of a handover was defined as the moment that the giver begins moving his/her hand to present the handover object to the receiver; the end of a handover was defined as the moment when the receiver has grasped the handover object. Using these defi-
nitions, observations were manually labelled in each trial between the start and end of a handover as positive (1); and those before or after a handover as negative (0). Thus, independent predictions can be made for each time step.

Since an objective of the classifier is to distinguish handovers from all other human motions, it is imperative that the negative (non-handover motion) data encompass a wide range of behaviours - not just those occurring before and after a handover. Thus, supplementary negative data was collected to introduce more diversity in the negative dataset to improve the generalization ability of the classifier. Giver motions in this dataset include more common actions such as walking past the receiver in a variety of directions and moving into sitting and standing positions, as well as uncommon actions such as performing jumping jacks and dance routines.

5.5.4 Test Set Generation

From the data set collected, the recordings from two participant pairs were reserved as a test set partition to be used to test the generalization of the trained SVMs and some of the supplementary negative data. The test set included approximately 25% of the total data. The rest of the data formed a training set used to train the SVMs.

5.5.5 Predictor Feature Selection

One of the disadvantages of using a supervised learning technique, such as the SVM, is the requirement to determine which features are relevant in producing a good recognition model. Often, in previous work, selection of features is performed in an ad hoc fashion [e.g., 112]. In contrast, this process is conducted systematically in the work presented here. To eliminate features that have a low predictive value in classifying positive and negative data (simplifying the model), an ensemble learning technique known as Breiman’s random forest algorithm [19] was used. The random forest algorithm is a well-known method for estimating predictor importances, that is, how well a single feature is able to discriminate between the target classes, reducing high-dimensional feature spaces into a core set of excellent predictors while also eliminating all irrelevant or redundant (correlated)
features [31]. Thus, this algorithm enables the ranking of the predictive ability of the 176 obtained features and provides insight into the importance of features for detecting handovers.

Random forests typically perform better in environments where learning models are diverse [71]. Bootstrap aggregating, or bagging, was used to grow the random forest as a method for increasing diversity between trees; i.e., weak learners were bootstrapped through random resampling of the data. Performance of bagged random forests relies on several parameters; two of the more influential ones being the size of the bootstrap samples, also known as the sampling fraction, and the number of weak learners in the forest. The sampling fraction is a critical parameter for the injection of diversity into weak learners, and the number of weak learners in the forest affects both computational efficiency and the robustness of the forest. The effect of both of these parameters can be measured using three error metrics: the generalization error, out-of-bag error and resubstitution loss. The generalization error is a measure of how well an algorithm performs on a hold-out partition reserved from the training data. The out-of-bag error is similar to the generalization error, but uses samples not seen in training as an independent test set. That is, for every example vector \((x_i,y_i)\) in the training data \(D\), select all bootstrapped subsets \(D_j\) that do not include \((x_i,y_i)\). Then, the example \((x_i,y_i)\) can be passed through all trees \(T_j\) trained on \(D_j\) for an aggregate prediction of that observation. This is carried out for all test observations and compared against the assigned labels. Lastly, resubstitution loss reflects the would-be test performance had the training set been the test set.

The effect of the sampling fraction on these error metrics was investigated by training multiple forests with 100 learners grown using varying sampling ratios. Similarly, the effect of the number of learners on the same measures were examined by training multiple forests of different sizes. Results using both Gini’s diversity index and information gain (also known as cross-entropy or maximum deviance reduction) split criterion were obtained. From Figures 5.2 and 5.3, it is apparent that all of the error metrics generally decrease with increased sampling fraction, and are minimally affected past 100 learners.

Based on these results, 500 weak learners were used to provide a sizable safety margin against any potential variations in the final model that could alter error
Figure 5.2: Charts depicting resubstitution loss, generalization error and out-of-bag error for Gini (left) and entropy/information gain (right) split criterion as a function of bootstrap sampling fraction.

Figure 5.3: Charts depicting resubstitution loss, generalization error and out-of-bag error for Gini (left) and entropy/information gain (right) split criterion as a function of number of learners.
rates. Each learner/tree was bootstrapped through random resampling of the data with replacement where the size of each bootstrap dataset was the same as the original, sampled dataset (sampling fraction = 1). A random subspace of features was chosen for each tree, the dimensionality of which was set to Breiman’s recommendation of the square root of the total feature space: \( \sqrt{p} \approx 11 \). Gini’s diversity index and information gain were both used as split criterion having three ensemble models trained for each.

Feature importance estimates were calculated by summing weighted decreases in Gini impurity as a result of splits on each feature and dividing by the number of branch nodes. Each ensemble produced a set of rankings of all features based on these importances, and the rankings for each feature was averaged amongst the six ensembles. The top 20 features were selected to form the input space for SVM training, not only to improve computational efficiency both in training and hyperparameter optimization, but also in online computation of features and prediction of new samples. In addition to these 20 base features, any right-hand positional (non-derivative) features not in the top 20 were included as well. Inclusion of the right-hand position features ensures that the classifier can still identify the handover event late if identification fails during the handover motion. In other words, the terminal stages of the handover, in which the giver’s arm is outstretched and idle, must be represented in the training data.

### 5.5.6 Hyperparameter Optimization

Four kernels were used for the SVMs: linear, quadratic, cubic and Radial Basis Function (RBF). The two tunable hyperparameters for each model are the regularization or penalty parameter, \( C \), and the kernel scale, \( \gamma \). The hyperparameter \( C \) operates as an adjustable tuning to control the degree of underfitting/overfitting of the data. For large values of \( C \), the optimization will choose a smaller-margin hyperplane in an attempt to correctly classify all training points. Conversely, a very small value of \( C \) will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points. The kernel scale hyperparameter controls the flexibility of the decision boundary. For example, with the RBF kernel, a larger \( \gamma \) leads to a decision boundary that forms a tight perimeter.
around a region where positive training data is clustered. The optimization objective is to find a hyperparameter pair that minimizes the cross-validation error.

Searching the parameter space requires the classifier to be trained using a time-intensive procedure at every iteration. Normally, a robust k-fold procedure is used for cross-validation optimization of the hyperparameters. However, if k-fold validation were used with, say, 10 folds, the classifier would need to be trained 10 times per iteration. Given that the training dataset has approximately 200,000 observations, a k-fold approach was deemed infeasible due to a lengthy execution time. Instead, a holdout cross-validation approach was used, while keeping in mind that the optimization results would not be as robust as in the case of using a k-fold scheme [18]. A holdout set randomly sampled with 90:10 split from the total training data was reserved prior to each optimization. Due to the size of the training data, a 10\% subset was deemed to be sufficiently large to capture the overall diversity of motions to be classified. Once optimal hyperparameters were obtained, five-fold cross-validation was carried out as a more robust evaluation of whether or not the optimization had indeed converged on an error minimum. Further reduction in computation time was achieved through the use of MATLAB’s built-in heuristic for automatically determining the kernel scale, effectively reducing the optimization problem into a univariate search.

For all of the kernels explored, it was assumed that the surface of the holdout error objective function was likely to be non-smooth and non-convex. Normally, a numerical method that could optimize non-linear problems such as Nelder-Mead would be used. However, due to the infeasible execution time it took to run the optimization on the large data set collected in this study, a hybrid of Golden Section Search (GSS) and Successive Parabolic Interpolation (SPI) was used instead. As this optimization method requires a unimodal objective function, the search was carried out over bracketed intervals of the search space to reduce the in-bracket function variance and capture potential convexity. The regularization parameter, C, was initially investigated over an initial range of (0,20]. This range was split into five intervals, with a minima search executed within each interval independently. Using this approach, five local minima were identified overall. The global minimum, or lowest function evaluation found within the search bracket, was taken as the kernels’ optimal hyperparameters. Significant downwards error
trends within the range [0.001,20] might suggest better minima exist and would
necessitate searching beyond $C = 20$.

5.6 Results

5.6.1 Training and Holdout Data

Table 5.2 shows the specifications of the resulting training and testing data (holdout partition). As mentioned in Section 5.5.4, approximately 75% of the total observations were used as training data and the remaining 25% were held out for testing. An approximate 50% split was maintained between positively and negatively labelled observations for both training and testing data sets.

Table 5.2: Training and testing data specifications for the automated detection of handovers. The total observations made, how many of those observations were labelled positively (indicating handover intent) and negatively (indicating non-handover), and respective percentages of total data collected are shown.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Observations</th>
<th>'+' observations</th>
<th>'+' (%)</th>
<th>'-' observations</th>
<th>'-' (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>205,365</td>
<td>93,407</td>
<td>45.5</td>
<td>111,958</td>
<td>54.5</td>
</tr>
<tr>
<td>Testing</td>
<td>50,845</td>
<td>27,722</td>
<td>54.5</td>
<td>23,123</td>
<td>45.5</td>
</tr>
<tr>
<td>Total</td>
<td>256,210</td>
<td>121,129</td>
<td>47.3</td>
<td>135,081</td>
<td>52.7</td>
</tr>
</tbody>
</table>

5.6.2 Predictor Feature Selection

The set of features chosen by the random forest method to form the input space for SVM training is shown in Table 5.3 in order of importance ranking. As expected from indications in prior work, right hand-, chest-, and forehead-related features score very highly. Y-axis kinematic measures are strongly represented, as are angular measures projected onto the XZ and XY planes. The receiver-based
kinematics class of features (e.g., velocity of the giver’s chest along the vector between the giver’s chest marker and the receiver frame origin) perform exceptionally well as predictive features, holding three of the top six places (ranks 1, 4 and 6) in ability to predictively classify data.
Table 5.3: The 22 features selected for use as the SVM input space to automatically detect handovers and their predictive ability rankings. The last two features (23 and 29) were not ranked in the top 20, but were included as part of the input space since they describe position of the right hand (the hand that was used to hand over objects in this experiment), such that the classifier can still identify the handover event later if identification fails during the handover motion.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Measure</th>
<th>Marker(s)</th>
<th>Feature</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Receiver-Based Kinematics</td>
<td>RHand</td>
<td>Velocity</td>
<td>RB_RHand_V</td>
</tr>
<tr>
<td>2</td>
<td>Linear Kinematics</td>
<td>Chest</td>
<td>Y- Component Velocity</td>
<td>LK_Chest_Vy</td>
</tr>
<tr>
<td>3</td>
<td>Linear Kinematics</td>
<td>RHand</td>
<td>Y- Component Velocity</td>
<td>LK_RHand_Vy</td>
</tr>
<tr>
<td>4</td>
<td>Receiver-Based Kinematics</td>
<td>Chest</td>
<td>Velocity</td>
<td>RB_Chest_V</td>
</tr>
<tr>
<td>5</td>
<td>Linear Kinematics</td>
<td>FHead</td>
<td>Y - Component Velocity</td>
<td>LK_FHead_Vy</td>
</tr>
<tr>
<td>6</td>
<td>Receiver-Based Kinematics</td>
<td>FHead</td>
<td>Velocity</td>
<td>RB_FHead_V</td>
</tr>
<tr>
<td>7</td>
<td>Linear Kinematics</td>
<td>RHand</td>
<td>Z - Component Velocity</td>
<td>LK_RHand_Vz</td>
</tr>
<tr>
<td>8</td>
<td>Angular Kinematics</td>
<td>RHand- RElbow</td>
<td>Angle w.r.t. XY Plane</td>
<td>AK_RHand-RElbow_xy</td>
</tr>
<tr>
<td>9</td>
<td>Linear Kinematics</td>
<td>FHead</td>
<td>Z - Component Velocity</td>
<td>LK_FHead_Vz</td>
</tr>
<tr>
<td>10</td>
<td>Angular Kinematics</td>
<td>RHand- RElbow</td>
<td>Angle w.r.t. XZ Plane</td>
<td>AK_RHand-RElbow_xz</td>
</tr>
<tr>
<td>11</td>
<td>Distance</td>
<td>RHand-RShoulder</td>
<td>Z - Component Distance</td>
<td>DI_RHand-RShoulder_z</td>
</tr>
<tr>
<td>12</td>
<td>Distance</td>
<td>RHand- Neck</td>
<td>Z - Component Distance</td>
<td>DI_RHand-Neck_z</td>
</tr>
<tr>
<td>13</td>
<td>Distance</td>
<td>RHand- RShoulder</td>
<td>Y - Component Distance</td>
<td>DI_RHand_RShoulder_y</td>
</tr>
<tr>
<td>14</td>
<td>Linear Kinematics</td>
<td>Chest</td>
<td>Z - Component Velocity</td>
<td>LK_Chest_Vz</td>
</tr>
<tr>
<td>15</td>
<td>Angular Kinematics</td>
<td>RElbow- RShoulder</td>
<td>Angle w.r.t. XZ Plane</td>
<td>AK_RElbow_RShoulder_xz</td>
</tr>
<tr>
<td>16</td>
<td>Distance</td>
<td>RHand- Neck</td>
<td>Y- Component Distance</td>
<td>DI_RHand_Neck_y</td>
</tr>
<tr>
<td>17</td>
<td>Angular Kinematics</td>
<td>LShoulder- Neck</td>
<td>Angular Velocity w.r.t. XZ Plane</td>
<td>AK_LShoulder_Neck_Vxz</td>
</tr>
<tr>
<td>18</td>
<td>Angular Kinematics</td>
<td>RElbow- RShoulder</td>
<td>Angle w.r.t. XY Plane</td>
<td>AK_RElbow-RShoulder_xy</td>
</tr>
<tr>
<td>19</td>
<td>Linear Kinematics</td>
<td>RHand</td>
<td>Z- Component Position</td>
<td>LK_RHand_Pz</td>
</tr>
<tr>
<td>20</td>
<td>Angular Kinematics</td>
<td>RElbow- RShoulder</td>
<td>Angular Velocity w.r.t. XZ Plane</td>
<td>AK_RElbow_RShoulder_Vxz</td>
</tr>
<tr>
<td>23</td>
<td>Linear Kinematics</td>
<td>RHand</td>
<td>Position Magnitude</td>
<td>LK_Rhand_Pmag</td>
</tr>
<tr>
<td>29</td>
<td>Linear Kinematics</td>
<td>RHand</td>
<td>Y- Component Position</td>
<td>LK_RHand_Vy</td>
</tr>
</tbody>
</table>
Axis component velocities and segment angles also score well. Four measures of distance between markers appear in the top 20, which relate the distance of the right hand to the neck base and right shoulder markers in the Y-(receiver forward) and Z- (vertical) axes, signifying that arm reach plays an important part in handover motions.

Overall, importance rankings, as seen in Figure 5.4, take a somewhat logarithmic shape. This suggests that the top end features are an excellent set of predictors, whereas the importance of other predictors rapidly deteriorates down the ranking. It would therefore offer little benefit to train on a larger number of features, especially considering the greater computational expense that would be incurred in so doing. The discriminatory ability of these features can be seen using a parallel coordinate plot (Figure 5.5). Here, the vertical axis represents the normalized coordinate value of observations for each feature listed along the horizontal axis. The features are listed in rank order from left to right along the horizontal axis. The plotted heavy lines link the average coordinate values for handover/non-handover for each feature, and the thin lines represent the quartile ranges surrounding the averages. These lines form threads that illustrate the ability of the features to separate positive and negative labelled data into disparate clusters. Using this format, strong clustering can be observed in the first six features where the quartile ranges are entirely separated. Proceeding to the right of the plot, features that are ranked lower demonstrate, as expected, less separation. This can be seen as further confirmation of the success of the bagged random forest approach for feature ranking for this problem.
Figure 5.4: Plot of the SVM feature space obtained from the bagged random forest arranged by importance scores in highest to lowest order. The horizontal black line delineates the top 20 features chosen to be included in the SVM feature space. Bars of similar colour represent features of the same type of measure.
Figure 5.5: Parallel coordinates plot of the input space determined by using a bagged random forest with 500 learners and sampling fraction = 1.0. The top 20 features are included, along with the two $RHand$ positional features not found in the top 20 but which are still included in the SVM input space. The thick blue dotted line represents the normalized mean value of the observations labelled negative (non-handover) within each feature space. Similarly, the thick green solid line represents the normalized mean value of the observations labelled positive (handover) within each feature space. Quartiles are shaded.
5.6.3 Hyperparameter Optimization

From Figure 5.6, the algorithm can be seen to converge on approximately convex regions of the search space. For the linear kernel optimization, there also appears to be instances of SPI degeneracy. In these cases, the algorithm has been able to proceed anyway, which may be due to the hybridization of the SPI and GSS algorithms. The search space plots also reveal that, as correctly predicted, all objective functions are non-smooth and, in fact, very noisy. For quadratic and cubic kernels, relatively large variations in cross-validation error exist between minute, fractional variations in the regularization parameter, C. This might suggest that the optimization would have benefited from a larger holdout set. However, the cross-validation error variations are relatively small as, overall, the objective function is extremely invariant (<1% holdout error) over the search space. Furthermore, the range of holdout error is far smaller than expected, suggesting a high accuracy and an exceptional level of separability. One might attribute the accuracy to overfitting; however, it is unlikely that this is the case as function evaluations are performed at C < 20 for all models. It is possible that small adjustments in C are leading to misclassification of only a handful of points, which, at this high level of accuracy, is sufficient to cause the apparent noisiness of the objective function. Regardless, the GSS/SPI hybrid optimization method has honed in on minima within each bracket.

Due to the invariance of the objective function, searching at C > 20 is unnecessary. While this invariance does decrease with increasing kernel complexity, even the RBF kernel’s error does not vary with more than a fraction of a percent. Furthermore, the maximum holdout accuracies of all models are all quite low. Satisfactory results have been found within the initial range, and the models will be more prone to overfitting at higher values.

Table 5.4 shows the values of the optimized hyperparameters for each kernel as well as the cross-validation errors for both holdout and five-fold procedures. The linear kernel’s global minimum is found in the first bracket, (0, 4], at C = 1.624. At this low level of regularization, the model will not penalize misclassification of the training data as strongly. The resultant low model complexity may lend itself to the overall generalisability of the final classifier. The quadratic, cubic and RBF kernels’ global minimum are found in the last bracket, [16, 20], at C = 17.528,
Figure 5.6: Kernel search spaces for optimization of the regularization parameter (C) for the handover classifying SVM. (a) Linear kernel; (b) quadratic kernel; (c) cubic kernel; (d) RBF kernel. The objective functions all appear to yield results that are localized to small holdout error and relatively invariant over the search range, invalidating the need to search beyond $C > 20$. 
17.655, and 19.416 respectively. This means that in comparison to the linear kernel, the remaining kernels will have a significantly higher level of regularization and model complexity. Nonetheless, with the knowledge of high holdout performance at these relatively low C values, it is apparent that there is an intuitive resistance to overfitting. Hence, although the cubic and RBF kernels demonstrated an inverse relationship between holdout error and regularization, these kernels would incur an unjustifiably higher risk of overfitting if searching at higher values of C.

Table 5.4: Optimized hyperparameter values and corresponding cross-validation error for the SVMs using various kernels.

<table>
<thead>
<tr>
<th>Model</th>
<th>C</th>
<th>γ</th>
<th>Holdout error (%)</th>
<th>Five-fold error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>1.624</td>
<td>2.624</td>
<td>3.57</td>
<td>4.0</td>
</tr>
<tr>
<td>Quadratic</td>
<td>17.528</td>
<td>2.644</td>
<td>1.48</td>
<td>2.0</td>
</tr>
<tr>
<td>Cubic</td>
<td>17.655</td>
<td>0.688</td>
<td>0.79</td>
<td>1.3</td>
</tr>
<tr>
<td>RBF</td>
<td>19.416</td>
<td>2.659</td>
<td>0.48</td>
<td>1.1</td>
</tr>
</tbody>
</table>

5.6.4 Confusion Matrix

The confusion matrix shown as Table 5.5 plots the resulting predicted classifications produced by the SVM with the four kernels tested against the actual classifications. The definitions for each of the metrics shown in the confusion matrix are displayed below in Equations 5.1 through 5.9.

\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{5.1}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5.2}
\]

\[
\text{False positive rate} = \frac{FP}{FP + TN} \tag{5.3}
\]

\[
\text{Positive predictive value} = \frac{TP}{TP + FP} \tag{5.4}
\]
Negative predictive value

\[\text{Negative predictive value} = \frac{TN}{TN + FN}\]  \hfill (5.5)

False omission rate

\[\text{False omission rate} = \frac{FN}{TN + FN}\]  \hfill (5.6)

False discovery rate

\[\text{False discovery rate} = \frac{FP}{TP + FP}\]  \hfill (5.7)

Negative predictive value

\[\text{Negative predictive value} = \frac{FN}{TP + FN}\]  \hfill (5.8)

Accuracy

\[\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}\]  \hfill (5.9)

where \(TN\) = true negative, \(FN\) = false negative, \(FP\) = false positive and \(TP\) = true positive.
Table 5.5: Confusion matrix for handover detection using a SVM with various kernels on motion capture data.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Actual condition</th>
<th>Prevalence: 55.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total Population:</strong> 50845</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Actual condition</strong></td>
<td><strong>0</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td>True Negative</td>
<td>LIN 43.3</td>
<td>LIN 1.9</td>
</tr>
<tr>
<td>False Negative</td>
<td>QUA 43.8</td>
<td>QUA 1.3</td>
</tr>
<tr>
<td>False Positive</td>
<td>CUB 43.9</td>
<td>CUB 4.3</td>
</tr>
<tr>
<td>True Positive</td>
<td>RBF 44.0</td>
<td>RBF 2.4</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>LIN 95.7</td>
<td>LIN 96.9</td>
</tr>
<tr>
<td>False Omission Rate</td>
<td>QUA 97.2</td>
<td>QUA 97.8</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>CUB 91.1</td>
<td>CUB 98.0</td>
</tr>
<tr>
<td>False Discovery Rate</td>
<td>RBF 94.8</td>
<td>RBF 98.2</td>
</tr>
<tr>
<td>Specificity</td>
<td>LIN 96.3</td>
<td>LIN 96.4</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>QUA 97.4</td>
<td>QUA 97.5</td>
</tr>
<tr>
<td>Accuracy</td>
<td>CUB 97.7</td>
<td>CUB 94.7</td>
</tr>
<tr>
<td>False Positive Rate</td>
<td>RBF 97.8</td>
<td>RBF 96.6</td>
</tr>
<tr>
<td>False Negative Rate</td>
<td>LIN 3.7</td>
<td>LIN 3.5</td>
</tr>
<tr>
<td></td>
<td>QUA 2.6</td>
<td>QUA 2.3</td>
</tr>
<tr>
<td></td>
<td>CUB 2.3</td>
<td>CUB 7.8</td>
</tr>
<tr>
<td></td>
<td>RBF 2.2</td>
<td>RBF 4.4</td>
</tr>
</tbody>
</table>
5.7 Discussion

5.7.1 Feature Selection

Prior work seems to indicate that being able to choose from a rich selection of kinematic features rather than simply using marker data (e.g., marker positions and velocities) improves classification ability [14, 79]. For example Begg and Kamruzzaman experienced poor accuracy (62.5%), sensitivity (66.7%), and specificity (58.3%) in classifying gait patterns when only baseline naïve kinematic features such as position and velocity were used in a RBF-based SVM [14]. However, these performance measures rise drastically (83.3% accuracy, 75.0% sensitivity, and 91.7% specificity) when many other kinematic features were considered as well, such as ankle and knee angles. Thus, one might expect that the exceptional results obtained in this study, as shown in Table 5.5 (94-96% accuracy, 92-97% sensitivity and 96-97% specificity), were obtained as a result of access to a richer kinematic feature set as opposed to using only position and velocity motion-capture information as features.

The features selected through the random forest algorithm are quite comparable to nonverbal cues investigated in other handover detection systems. More than half of the features selected relate to the kinematics of the giver’s right arm and hand (the hand that was used to hand over objects). Four of the top six features relate to the approach of the giver towards the receiver (e.g., chest and forehead positions and motions) and two features relate to the vertical motion of the giver’s body. The importance of these features in handover detection supports prior work that has either observed [13, 54] or used [36, 85] similar features. Thus, this work provides further confirmation that features deemed important in previous studies are indeed useful for handover detection through a systematic method of kinematic feature selection. However, in previous studies only the location and velocity of the hand, or joint angle relations between elbow, shoulder and hand, were observed/used. The feature set presented here indicates that several more features relating to the Cartesian and joint angle features of the giver’s arm involved in the handover could be used for the detection of handovers (e.g., component distances between joint locations and joint angle velocities) than those traditionally examined in prior work.
It is interesting that a left-sided feature, namely, the angular velocity of the left-shoulder to neck segment with respect to the XZ plane - was also found to be in the top 20 features. Both video footage and motion-capture data from the experiment were reviewed to investigate why this feature was included. From this review, it was found that in the majority of handover motions, the giver twisted their body about their center vertical axis such that the right shoulder rotated towards the receiver, extending the reach of the right hand and causing the left shoulder to rotate away from the receiver. Although the observance of this rotation seems natural, no reference to it as a characteristic cue in previous handover studies was found. One explanation stems from proxemics and the experimental set up: as most givers and receivers had no relationship outside the experiment, givers were more likely to approach only to within a social distance (>1.25m) [45, 121]. Thus, to facilitate the handover at that distance, the giver rotates their body to extend their reach. An alternative, though equally likely, explanation for this behaviour is that the giver performs the handover in a way that is believed to be energy-efficient given the situation, in the same way that a person lying on a couch is more likely to try to reach for a remote control with an outstretched arm rather than walk to retrieve it. Whatever the case may be, the results of this study suggest that this rotation should not be overlooked as a feature in handover detection.

None of the features in the input space reflect the proximity of the giver to the receiver, except for the Y- component position of the right hand, which was later added to the feature set. This is unusual as several studies have observed patterns in distances between giver and receivers during handover [13, 84, 115]. However, Strabala et al. also mentions that proxemic features are not used in their classifier [113]. A review of scatter plots of specific marker positions within the training set seems to explain this absence: clusters of positive and negative observations seem to perfectly overlap each other, leading to positional features having poor discriminatory ability by the random forest algorithm. This may be an unintended consequence of collecting data from a laboratory environment where the participants’ initial proximal positions are similar to the positions seen during handover. Strabala et al. offer another explanation: that givers start handover motions independently of the distance from the receiver. Regardless, it could be argued that the absence of proxemic features reduces the rate of false positives of handover detec-
tion occurring due to two people standing in close proximity, without handing over objects. More investigation is required.

Given the exceptional performance of the classifier on these selected features and that the importance of these features falls off logarithmically (Figure 5.4), and that the number of features used in the classifier was arbitrary, it may be possible to further reduce the number of features used by the SVM classifier. Instead of selecting the top 20 features, it may be possible to select a smaller number of features, e.g., the top ten features or a subset of the top 20 that are the most accurate or efficient to measure given current technologies, without significantly degrading the classifier’s performance. This would not only be beneficial in reducing the complexity and execution time of training and running the classifier models, but also reduce computational complexity if these models were to be run online.

Overall, this study demonstrated that the use of the bagged random forest method for the automatic ranking and selection of predictive features indicative of handover motions provides intuitive and useful information regarding common nonverbal cues that can be used to anticipate object handover motions from the perspective of the receiver. The high accuracy of the SVM models using these features provides additional evidence for the importance and discriminatory ability of these automatically-selected kinematic features, validating the use of the bagged random forest approach to feature selection. Furthermore, it is reassuring that an autonomous algorithm was able to select features that were highlighted by previous works as potentially important nonverbal cues in handovers.

5.7.2 Model Verification

Test data performance from all SVM models far exceeds expectations. All models have specificity, sensitivity, precision (positive predictive value), negative predictive value, and overall accuracy in excess of 90%. Linear and quadratic kernels slightly outperformed the cubic and RBF kernels both in overall accuracy and sensitivity, though only marginally. Otherwise, all kernels perform similarly on the holdout test set. This seems to signify that the hyperparameters selected for each kernel was able to balance the trade-offs between overfitting and underfitting the data, allowing the models to generalize well to the holdout test set. A review of
the classification results from the time series suggests that approximately half of the classification errors occurred near the start and end of handovers at locations where labelled classifications transition from nonhandover to handover. In effect, the SVM classifiers tend to detect handovers slightly earlier or later than how the data is labelled, which is to be expected due to imperfections in the labelling process. With a 30 Hz sampling frequency, this may represent a few milliseconds time difference in detection.

Although results suggest that using any of the kernels should produce adequate handover detection, it may be more robust for robots to use the RBF kernel rather than any of the polynomial-based kernels to limit false positives. To explain, consider the case of using the linear kernel: a linear discriminant function can be thought of as a weighted sum of normalized features and, thus, any abnormally large value for any one of the features could cause the model to falsely detect a handover. For example, running towards a robot equipped with a linear discriminant handover detector will most likely produce a detection due to abnormally high chest and right hand velocities in the Y-axis, thus causing the linear detector to declare a positive detection when, in fact, no handover occurred. Similar problems may exist for higher order polynomial kernels as well, though a nonlinear kernel (such as RBF) or additional rules for validating data could alleviate these issues.

The results from the testing of these classifiers show that classification performance exceeds that of the classifiers developed by Strabala et al. which had an average classification accuracy of 80% (peaking at 89%) [112]. It is evident from this comparison that the application of SVMs on kinematic data is definitely a promising method for the detection of handovers. Our intuition suggests that other machine-learning approaches, e.g., neural networks and random forests, may also be similarly successful in performing classification on kinematic data.

5.7.3 Extendibility of Method to Other Trackers

Tracking human motion with depth cameras such as the Intel RealSense, PrimeSense, and Microsoft Kinect is a more attractive option for developing human-robot interactive systems than traditional motion-capture setups, since these sensors:
• are inexpensive compared to the cost of a motion-capture setup;
• can be easily mounted to a mobile robot that is not restricted to travel within a confined space (as opposed to having to stay within a motion-capture tracking area);
• do not require users to wear garments with retroreflective markers; and
• require only one camera sensor rather than the handful need by motion-capture systems.

The widespread availability of off-the-shelf body tracking sensors such as the Intel RealSense or Microsoft Kinect version 2 makes the implementation of machine-learning techniques based on kinematic data more straightforward than previous techniques, especially those based upon image processing used in several prior works relating to gesture detection and handovers [34, 112]. To provide insight into the feasibility of this method for such sensors, a pilot study was performed using a Microsoft Kinect version 2 sensor for online detection of handovers. For this study, an experimental setup and participant task similar to that described in Section 4.4.4 was used; however, a Kinect was situated slightly above eye-height in front of the receiver to track the giver’s upper body. As the Kinect’s skeleton tracking uses similar key body points (Figure 5.7) compared to the motion-capture marker placements used in the main study (Figure 4.4), the same selection of predictor features as found in Section 5.5.5 were used - features were computed directly from the Kinect body points that were within the vicinity of the original markers.

Because of the slight differences in tracking, the previously obtained SVM models were not reused; instead, a new SVM was trained using an RBF kernel (for reasons presented in Section 5.7.2) with a much smaller dataset consisting of 50 handovers from a single giver (15,312 observations, 55% negative, 45% positive). The LibSVM library [30] was used to perform model training. This model was then applied to an additional set of 25 handovers collected from the same participant for handover detection. Detection was performed online (as opposed to offline processing performed in the main study) at the Kinect’s frame rate of 30 Hz to determine the feasibility of using this handover detection method for human-robot
handovers. Similar to the post-processing that was performed on marker data in the motion-capture data for the main study (Section 5.5.1), missing tracked points were inferred automatically within each frame by the Kinect’s skeleton tracker and the data was filtered using the same 4th-order Butterworth filter prior to feature generation.

The results are shown in Table 5.6. It is important to remember that this is a pilot study that uses training and testing data from the same participant: a classifier that incorporates data from multiple participants may strengthen or weaken the results as seen here. Thus, only trends as seen from the data can be are discussed. In the resulting confusion matrix obtained from the pilot study, it can be seen that the statistics are similar to the results in Table 5.5: accuracy, sensitivity, and specificity
are all above 80%, though marginally lower than what is observed from the RBF kernel performance in the main study. This slight performance decrease was to be expected due to the increased noise from the Kinect needing to fit a skeleton model to time-of-flight depth data.
**Table 5.6:** Pilot study confusion matrix for handover detection using a SVM with RBF kernel applied to features generated via a Kinect version 2 sensor.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Actual Condition</th>
<th>Prevalence: 48.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total Population: 7312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>True Negative 49.1</td>
<td>False Negative 2.0</td>
</tr>
<tr>
<td></td>
<td>False Positive 5.7</td>
<td>True Positive 43.2</td>
</tr>
<tr>
<td>1</td>
<td>Specificity 89.6</td>
<td>Sensitivity 95.5</td>
</tr>
<tr>
<td></td>
<td>False Positive Rate 10.4</td>
<td>False Negative Rate 4.5</td>
</tr>
<tr>
<td></td>
<td>Accuracy 92.3</td>
<td></td>
</tr>
</tbody>
</table>
Although the preliminary results indicate good performance, there are some issues that may hamper the method’s applicability to the Kinect and similar sensors. One concern (though not observed from the pilot study) is that objects may occlude hands or arms or cause poor fitting of the skeleton model. This, in turn, may lead to imperfect tracking and representation of joint features, causing poor prediction performance. This issue can be remedied with advanced object detection and segmentation methods, such as those employed by Schmidt et al. [103]. Detection accuracy and speed may also be of concern with a larger pool of participants due to the noisy nature of data obtained from such devices. There are also limitations due to assumptions that were imported in terms of experimental control. One such limitation relates to inherent constraints placed on participants due to the experimental setup: givers initiated the handover directly in front of the receiver and in view of the camera; as opposed to having participants moving fluidly around the room and performing handovers in less constrained locations and orientations. Needless to say, handover gestures occurring out of view of the camera would not be detected. Also, if participants are positioned with the giver standing side-by-side or off to the side of the receiver as opposed to face-to-face (which was the default position used in the experimental setup), such handovers will likely not be detected as the SVM of this work has not be trained to recognize these handover configurations and the Kinect (as well as similar body tracking cameras) require an unobstructed view of the upper body to track a person’s limbs. Being able to have the system recognize new handover configurations is a matter of retraining SVM with a wider range of participant positions during handover. However, the latter issue of having the Kinect ‘see’ the receiver and the handover gesture is a shortcoming of using the Kinect. Given that the field-of-view of the Kinect is much more limited compared to the the motion capture setup used in the main experiment, optimal placement of the sensor becomes an important question. For a humanoid robot, such as the one on a PR2 robot (Willow Garage Inc., Menlo Park, California, USA), placement of the sensor on the articulated anthropomorphic head (which happens to be the default placement of the sensor on the PR2 as recommended by Willow Garage) may mitigate this issue. Since the handover detection method is receiver-centric in that the features used by the SVM are measured with respect to a coordinate frame
centered at the receiver, having the Kinect mounted on a robot head makes sense for many reasons:

- Articulation of the head can allow the Kinect to sense a wider field in space.

- The receiver coordinate frame used as a frame of reference for feature calculations rotate along with the head. Thus, as long as a human giver approaches and initiates a handover in the direction of the gaze of the robot head, the handover should be detectable by the SVM.

- Users can obtain a sense (via comparison of expectations during human-human handovers) of whether a handover gesture is detectable by the robot based on the position of the head that is consistent with the sensor’s field-of-view limitations - e.g., if the robot head is turned away from the giver, the giver may recognize that the robot may not be able to see their handover gesture and re-approach the robot from a direction in line with the robot’s gaze.

For non-humanoid robots, specifically robot arms, placement may be trickier. One solution may be to arbitrarily affix the sensor to a stationary feature, and constrain the direction of handover interaction. Alternatively, an array of Kinetics mounted on the base of the robot to obtain greater coverage of the field.

Overall, however, the results of the pilot study suggests that the online application of the handover detection method on data streamed from the Kinect, and perhaps other skeleton tracking sensors, is viable and can be considered as an option for robots to detect handover gestures. Additionally, there is also evidence that the feature set selected for recognizing handover gestures is able to extend readily to the Kinect version 2, despite slight differences in the skeleton models that are tracked by the motion-capture system and Kinect version 2. This implies that no recoding and/or reselection of features is required for data gathered from the Kinect version 2 (and possibly similar sensors such as the PrimeSense or Intel RealSense) to adequately classify handover gestures.
5.8 Conclusions

In this study, multiple handover-detecting SVM models that exceeded the objective of >80% classification accuracy (with a peak of 97.5%) were developed upon human kinematic data. Having been successful in using this method for handover detection, the following statements appear to be supported by the results:

- Kinematic data obtained directly from motion-capture devices are able to offer information on the intent to hand over.

- Using machine learning to both automatically select features that may offer high discriminative ability and perform classification on kinematic data offers great promise in the area of handover detection, as well as gesture detection.

Although the SVM algorithm and features used are certainly not novel, the systematic method for automatically determining kinematic features - confirming the importance of some features found in prior work - and the application of these features to handover detection using the SVM offers a new method for advancing shared object manipulation in human-robot interaction, where prior methods include image analysis and manual coding of time-series data. With this result, it is feasible to imagine that a robot equipped with an SVM classifier and skeleton tracking sensor can accurately recognize when a human is intending to pass an object to it and prepare a trajectory to receive that object. Although the Kinect version 2 has a frame rate 10× slower than the Vicon system and is susceptible to noise when capturing joint and limb data, the use of down-sampled motion-capture data and filtering in the analysis has indicated that detection is possible under similar conditions. This hypothesis is further supported by a preliminary pilot study in which handover detection is performed online with data from a Kinect version 2 sensor. Thus, this work suggests that machine learning on kinematic data offers a highly promising approach to allow robots to detect handovers, directly addressing the second research question posed in this thesis.

The ability for a robot helper to accurately detect handovers has a variety of implications for the degree in which robots will be able to autonomously and appropriately respond to users within the home or workspace. Most of these envi-
ronments have tasks that rely on the dexterity of humans and cannot be completely surrendered to automation. However, having robots that can assist persons by handing over or receiving tools and supplies much like a nurse assists a surgeon during an operation could allow robots to be more valuable in these environments. For these application areas, a handover detection system would be an invaluable addition to a robot’s toolbox of abilities. Thus, the work presented in this paper should be of great interest to robot designers, integrators, and users alike.

5.8.1 Limitations

Experiments and data collection for this work was performed in a controlled laboratory setting where the handover procedure was structured (e.g., handovers were right-handed, handovers occurred with participants facing each other). Additionally, the data used for this study was collected under the pretext of another study. As such, some of the procedures used to collect this data were not optimal for this work. Both of these factors may have influenced handover behaviours. The number of participants recruited for this study was also limited. Another key limitation of this study lies in the classifier: the classifier, by design, is focused solely on discrimination and may ignore kinematic features that are redundant, common to both positive and negative training data, and have little or no perceived additive value in discriminating between negative and positive handovers. For example, it is quite possible that there is a feature that, when used alone, is very good at discriminating between positive and negative examples, but is ultimately eliminated by the classifier as it produces no additional benefit compared to the mosaic of features already included in the classifier. Hence, there may be features either explored or beyond the scope of this work that could be beneficial in detecting handovers, but have been ignored.

5.8.2 Future Work

As the data was processed offline, the application of SVMs or other supervised machine-learning methods for online detection of handovers should be explored. Current trends in machine learning lean towards unsupervised learning as a method for gesture recognition, and current work in this area has shown some promise [86].
Such data-driven approaches rely only on raw data and do not require any pre-processing or test sets to teach classification, in contrast to what was performed in this work. Although good preliminary results were observed from transplanting the SVM from a motion-capture system to a Kinect, unsupervised learning offers greater flexibility in deployment to other sensors, where its algorithms can adapt to different data streams to classify gestures. In this sense, using skeleton tracking is not necessary with unsupervised learning, and raw data from depth cameras could be used to drive classification instead. However, because unsupervised learning self-clusters data into classifications rather than forcing data into pre-defined ones, a generic gesture clustering system would most likely be developed rather than a handover detector. Additionally, the association of a learned gesture with an appropriate reaction to that gesture must be learned as well, increasing the complexity of the system. Regardless, it would be interesting to see if unsupervised learning methods might similarly excel at recognizing users’ intent to hand over objects.

A more extensive study of using a body tracking system that is more likely to be used on a robot (e.g., PrimeSense, Intel RealSense, and Kinect sensors) rather than using scrubbed data from a Vicon system would be useful. Pilot data on one participant presented in this work indicates that the feature-based SVM approach to detecting handovers is easily transferable to the Kinect version 2 sensor with good results, though a more expansive user study would definitively confirm these findings.

The SVM classifiers developed are specific to right-handed handovers alone. There are several methods for remedying this issue, however, which include the use of a more diverse data set that includes left-handed handovers or a mirroring trick (e.g., running two classifiers simultaneously with one using giver data that is mirrored about the YZ plane and taking the maximum of the two hypotheses generated).

Due to the high accuracy obtained with pure handover prediction alone, there is potential for classification of more specific characteristics of the handovers. A significant diversity of objects was used in the handover experiment, and so it may be possible to classify object properties. For instance, the weight or fragility of the handover object could be ascertained through features such as handover speed, joint configuration, or one- versus two-handed grasps. Here, it might be most
telling to also train on the object pickup time interval (e.g., when picking up more fragile objects, givers may spend more time adjusting grasp before performing a handover). Successfully classifying these properties would significantly simplify the regulation of the robot’s receiving behaviour, as appropriate adjustments can be made towards the expected weight and fragility.
Chapter 6

Evaluating Social Perception of Human-to-Robot Handovers

Having demonstrated a method that enables robots to react to handover intent within human-to-robot handovers in the previous chapter, here, an exploration of the design space is initiated for robot receivers in terms of nonverbal gestures and cues. In the same vein of research presented in Chapter 3 where robot gaze was examined as a nonverbal cue that can affect factors of robot-to-human handovers, the work presented here and in Chapter 7 continues to investigate how certain nonverbal cues may affect factors of human-to-robot handovers. Particularly, this chapter examines how robot nonverbal cues such as initial pose, grasp method, and retraction speed of the robot before, during and after the handover, respectively, affects how people qualitatively perceive the robot. Being able to determine how nonverbal cues change how users socially perceive robots may help determine how to improve the handover interaction. To evaluate user perceptions, a recently developed psychometric tool called the ROSAS, developed by Carpinella et al. [25], is used.

A portion of this work presented in this chapter was disseminated at the Human-Robot Interaction in Collaborative Manufacturing Environments (HRI-CME) workshop of the 2015 IEEE/RSJ IROS held on September 24, 2017 in Vancouver, BC, Canada. The title of the submission which I co-authored was “Validation of the Robotic Social Attributes Scale for Human-Robot Interaction through a Human-
to-Robot Handover Use Case”. The contents of this chapter were presented at the 13th ACM/IEEE International Conference on Human-Robot Interaction held from March 5-8, 2018 in Chicago, Illinois, USA. A version of the manuscript included in the proceedings entitled “Evaluating Social Perception of Human-to-Robot Handovers using the Robotic Social Attributes Scale (ROSAS)” is reproduced in this thesis with only minor modifications. Supplementary materials relating to the experiment conducted in this work can be found in Appendix C.

6.1 Introduction

The ability of robots to safely and effectively handover objects is a crucial capability for collaborative human robot interaction. It is a skill that will allow robots to be increasingly useful in contexts such as manufacturing and assistive care. In particular, handovers between robot and human agents have previously been well studied. As robots are often tasked with delivering objects, most work places the robot in the giver role. In this work, focus is given to the reverse interaction, namely, handovers from humans to robots.

Previous studies have shown that nonverbal cues such as gaze and kinematics have significantly affected fluency, legibility and efficiency of robot-to-human handovers [3, 21, 22, 113]. Given the importance that nonverbal cues play in such interactions, it can be argued that similar attention should be given to nonverbal cues displayed by the robot during human-to-robot handovers. Yet, this topic is relatively unexplored by prior work. Consequently, an exploration of the design space is necessary to determine how best to facilitate human-to-robot handovers. As a first step, this work investigates user perceptions of the robot during this interaction.

The goal of this work is to gain a deeper understanding of robots in a receiver role and how factors influence users’ behaviours and perception. The study presented here examines how human collaborators perceive their robotic counterparts from a social perspective during object handovers. Specifically, this work explores how changing conditions of how the robot receives an object may change user opinion of the robot. Such a user-driven approach to exploring the design space allows for identification of facets of the interaction which serve as key elements.
in an interaction from those that do not, and hence, may inform the selection and
development of robot behaviours which may be better suited for HRIS. Thus, this
examination of how users socially view and evaluate collaborative robots may lead
to a greater insight of what features, characteristics, and/or behaviours of robots
can drive more engaging, efficient and fluent handovers, and more broadly, HRIS.
For this exploration, the Robotic Social Attributes Scale (ROSAS) - recently devel-
oped by Carpinella et al. [25] - is used as a tool for measuring user perception of
the robot.

6.2 Background

6.2.1 Handovers

Prior work studying handovers has mainly focused on human-to-human and
robot-to-human handovers. Among these studies, there seems to be no consensus
on which set of factors are important in determining how handovers are carried
out; rather, a survey of the literature indicates that a multitude of unique factors
affect how a handover is carried out by participants. Many studies considered how
seemingly inconspicuous nonverbal cues can play an important role in coordinating
and directing handovers [3, 13, 22, 26, 53, 77, 88, 112, 123]. For example, multiple
studies have found that gaze and eye contact for both humans and anthropomorphic
robots can affect timing and coordination of handovers [3, 88, 112, 123]. Another
stream of work has examined how grip and load forces plays an important part
in allowing givers and receivers negotiate handovers leading to insight into the
roles participants assume within a handover interaction [26, 63]. Other studied
factors include arm kinematics and movement timing [3, 22], proxemics [13, 68]
and handover object orientation [6, 29].

Since the number of factors within the design space for human-to-human and
robot-to-human handovers appears vast, it can be expected that the design space
for human-to-robot handovers would be no different, although largely unexplored.
The work presented here begins by delving into factors that are expected to af-
fect perceptual judgments of the robot receiver, measured using the Robotic Social
Attributes Scale.
Robotic Social Attributes Scale (ROSAS)

In prior work, many studies of HRI have employed the Godspeed scale developed by Bartneck et al. [12]. The Godspeed scale features five dimensions for rating robots: anthropomorphism (human-like vs machine-like), animacy (how life-like the robot appears or behaves), likeability (how friendly a robot seems), perceived intelligence, and perceived safety. However, despite its widespread appeal, Ho and MacDorman and Carpinella et al. have found shortcomings to the scale including: lack of empirical work examining its psychometric properties, occurrences where scale items are confounded with positive and negative affect, situations where items do not correspond to the underlying constructs they are meant to measure, high correlations between constructs, and multidimensionality of some item pairings [25, 49].

Thus, in an effort to provide a more valid scale, Carpinella et al. developed the Robotic Social Attributes Scale (ROSAS) which attempts to address these issues through exploratory factor analyses and empirical validation. The ROSAS is a social psychometric instrument aimed towards measuring social perception and judgments of robots across multiple contexts and robotic platforms [25]. The development of the ROSAS is based upon the Godspeed scale and claims to improve cohesiveness, eliminate unnecessary dimensions through factor analysis, and not be tethered to specific types or models of robots. The scale measures three underlying robotic attributes - competence, warmth, and discomfort using 18 items which are shown in Table 6.1. While the scale borrows the competence and warmth factors from more standard psychometric instruments used in social psychology measuring social perception [38], work by [25] shows that evaluations of robots are somewhat more complex, employing the third, discomfort factor that is additionally measured by the ROSAS. The scale was validated by Carpinella et al. via a study which had participants evaluate gendered human, robot, and blended human-robot faces shown on a screen. In contrast, this work proposes to use the ROSAS to evaluate a physical HRI.

The ROSAS has been chosen for this work as it provides a empirically validated method of measuring how users perceive their robotic counterpart. Additionally, as the ROSAS shares the competence and warmth dimensions with measures of social
Table 6.1: Table of ROSAS items categorized by the factors competence, warmth and discomfort.

<table>
<thead>
<tr>
<th>Competence</th>
<th>Warmth</th>
<th>Discomfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable</td>
<td>Organic</td>
<td>Awkward</td>
</tr>
<tr>
<td>Competent</td>
<td>Sociable</td>
<td>Scary</td>
</tr>
<tr>
<td>Knowledgeable</td>
<td>Emotional</td>
<td>Strange</td>
</tr>
<tr>
<td>Interactive</td>
<td>Compassionate</td>
<td>Awful</td>
</tr>
<tr>
<td>Responsive</td>
<td>Happy</td>
<td>Dangerous</td>
</tr>
<tr>
<td>Capable</td>
<td>Feeling</td>
<td>Aggressive</td>
</tr>
</tbody>
</table>

perception of people, it allows for intuitive comparisons and extrapolation of how the robot may be matched against a human in terms of these dimensions.

6.3 Experimental Design

As this work is mainly an exploration of design space for robot receiving during handovers, a small number of variables were selected from a potentially large pool to test to limit the scope of the study.

Several criteria were applied for selecting factors to examine. Factors selected:

- affect the chronological beginning, middle, and end of the handover (to allow multiple factors to be tested in a condition);
- were previously studied in human-human interactions, the results of which could be used as benchmarks and/or for drawing comparisons;
- are impartial to participants’ dominant/sub-dominant handedness (to limit complexity of implementation of the experimental setup); and
- contain levels achievable by the experimental setup, as shown in Figure 6.1

Three factors which emerged as variables for robot receiving during handovers during a pilot study are systematically tested in this study: initial position of the arm prior to handover, grasp type, and retraction speed following handover. To
constrain the length of time needed to run the experiment for each participant, the number of levels per factor were limited to two.
Figure 6.1: Experimental setup for the human-to-robot handover experiment. Diagram shows both the up and down initial arm positions tested as conditions in the experiment.
6.3.1 Initial Position of the Arm Before Handover (Down and Up)

For this factor, the initial arm position of the robot displayed to the giver is modified prior to handover. Two positions are used which are labelled up and down. Both initial positions are shown in Figure 6.1 The initial arm position was chosen to be examined as a factor since it is expected that differences in pose of the robot may affect giver behaviour when they are reaching out to indicate where and when a handover takes place. For example, the up position could convey the robot is awaiting the handover object, whereas the down position might suggest that the robot has not yet recognized the givers intent. They also present slightly different initial spacing between the robot end effector and user, which may affect where the handover takes place as indicated by Huber et al. in [54] and Basili et al. in [13]. Examination of differences in handover locations between both of these levels may provide insight into how users behave, in terms of proxemics, to a disembodied robot arm verses human/humanoid agents.

6.3.2 Grasp Type During Handover (Quick and Mating)

Motivated by prior studies on haptic negotiation in Human-Computer Interaction (HCI) which suggest that dynamic interactions are able to change how 'personal' and 'human-like' an interaction is [44, 94], robot grasping is examined as another factor in this work. Gripper design and grasping is still an active area of research. Much of this work tries to solve the problem of matching the speed, smoothness, dexterity and conformity of the human grasp. Current state-of-the-art grasping methods either carefully plan feasible grasps and execute them slowly, or applies brute force to 'robotically' grasp without delicacy of human touch. Rather than focus on object grasping, a simple co-planar interface (electromagnet) is used to allow for emulation of both extremes. With this grasp method, speed and brute force can both be achieved by turning on the magnet which creates sudden impulses due to minute misalignments. Alternatively, misalignments can be slowly accommodated for to create a smooth yet slow contact. In the quick grasp, the robot moves its electromagnetic end effector to within 1 cm distally from the cap of the baton during a handover. As soon as the 1 cm threshold is met, the electromagnet is activated and draws in the baton. In the mating grasp, the robot deliberately
moves all the way into contact with the baton. Then, based on measurements of an ATI Mini45 Force/Torque sensor (ATI Industrial Automation, Apex, North Carolina, USA) located in series with the electromagnet at the robot end effector (see Figure 6.1), it further adjusts its orientation to achieve flush contact. Only when the electromagnet is coplanar with the baton’s cap is it activated. This behaviour allows the robot to ensure stable contact and thus safety of the object during handover before retracting. A flowchart of how these grasping behaviours are carried out can be found in Figure 6.2.
Figure 6.2: Flowchart of *quick* (top row) and *mating* grasp types. The *quick* grasp pulls in the baton magnetically while the *mating* grasp establishes coplanar contact, gently pressing against the baton before activating the magnet.
6.3.3 Retraction Speed Following Handover (Slow and Fast)

Retraction speed was selected as a factor for examination as prior work has shown that a robot’s speed of movement seems to play a significant role in how human observers and collaborators subjectively perceive the robot \cite{95, 101, 124}. For example, in an experiment conducted by Zoghbi et al., they found that fast robot motions were correlated with increased user arousal and decreased valence during self-reports of affect \cite{124}. Thus, in this work, retraction speed following object handover is hypothesized to affect how users perceive the robot in terms of the ROSAS measures of warmth and discomfort - e.g., slow retraction speed may be rated as higher warmth and lower discomfort as opposed to higher speed which may lead to less warmth and greater discomfort. The slow setting was set to 10 cm/s, whereas the fast setting was set at 20 cm/s. These settings were designed to emulate a gentle tug and a firm yank.

6.3.4 Conditions

A 2x2x2 experiment design was used to test these factors, which form 8 conditions, as shown in Table 6.2, which were counterbalanced between participants using a Latin square design to prevent carry-over effects. These factors were not only examined to see how they affect user perception of the robot’s attributes, but also to study how they affected proxemics and kinodynamics of the handover interaction. For example, it was hypothesized that examining initial arm position could help determine how people approach and direct handover gestures to a disembodied robot arm and how these gestures compare to human receivers studied in prior work \cite{13, 96}; retraction speed and grasp type were selected to research the force/torque interaction between the giver and receiver and to establish what dynamic negotiations occur during human-to-robot handovers. However, investigations of the impact of these factors on physical characteristics of handovers is beyond the scope of this study and is left as future work.
Table 6.2: Table of experimental conditions for the human-to-robot handover study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Arm Position</th>
<th>Grasp Type</th>
<th>Retraction Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>down</td>
<td>quick</td>
<td>slow</td>
</tr>
<tr>
<td>B</td>
<td>up</td>
<td>quick</td>
<td>slow</td>
</tr>
<tr>
<td>C</td>
<td>down</td>
<td>quick</td>
<td>fast</td>
</tr>
<tr>
<td>D</td>
<td>up</td>
<td>quick</td>
<td>fast</td>
</tr>
<tr>
<td>E</td>
<td>down</td>
<td>mating</td>
<td>slow</td>
</tr>
<tr>
<td>F</td>
<td>up</td>
<td>mating</td>
<td>slow</td>
</tr>
<tr>
<td>G</td>
<td>down</td>
<td>mating</td>
<td>fast</td>
</tr>
<tr>
<td>H</td>
<td>up</td>
<td>mating</td>
<td>fast</td>
</tr>
</tbody>
</table>

6.4 Experimental Setup

6.4.1 System

A KUKA LBR iiwa 7 R800 robot (KUKA, Augsburg, Germany) was used in this study to receive objects from participants. The robot was mounted as shown in Figure 6.1, 135 cm above ground level and fitted with a simple electromagnetic gripper. When activated, the gripper allowed the robot to securely grasp a handover baton via coplanar interfacing with a ferromagnetic cap mounted to the top end of the baton.

A set of 12 OptiTrack Flex 13 motion capture cameras (NaturalPoint, Corvallis, Oregon, USA) were used to track objects within an approximately 3x3m space. Each tracked object uses a unique constellation of retroreflective markers. The user’s hand, handover object, and robot end effector were tracked. The Flex 13 cameras have a frame rate of 120 frames per second with an average latency of 8.33ms (as reported by OptiTrack’s Motive software). Position and orientation tracking data of each object were transmitted via UDP to a second computer controlling the robot’s behaviour.

For this system, a handover model which stipulated that the robot receiver reacts to the giver is used. Thus, in the study, participants initiated the handover by
holding out the baton towards the robot, similar to how handovers have been initiated in previous studies [96]. The robot checked to see if the baton is in its reachable workspace; if so, the robot proceeded to move to grasp the object from its initial position. Once certain grasp conditions were met (determined by the grasp condition), the robot activated the electromagnet and began retracting the arm and baton by 10cm, before moving into the arm down position (see Figure 6.1). If, at any point during the retraction and movement to the arm down position, the system detected that the baton was not being grasped (e.g., the giver did not release the baton and overcame the electromagnet), the robot immediately returned to the baton to reattempt grasping.

6.4.2 Participants

This study was reviewed and approved by the Disney Research Institutional Review Board. A priori power analyses were conducted to determine the sample size required for this study. With $\alpha = .05$, a sample size of 20 was needed to detect a moderate effect size ($\eta^2_{\text{partial}} = .13$) with 90% power ($1 - \beta$) [32]. Recruitment was performed within Walt Disney Imagineering Advanced Development and Disney Research Los Angeles (DRLA). Twenty-two participants (11 females, 11 males), aged 22-52 years [$M=30.32, SD=8.12$] were recruited in total. All participants provided their informed consent prior to the experiment using the form shown in Section C.1; they were notified that their participation was voluntary, and they were allowed to withdraw from the experiment at any time. Additionally, permission was obtained from all participants to record both video and motion capture data from the experiment. No reward was given for participation in this study.

6.4.3 Participant Task

Participants were asked to provide consent to participate in the study and acknowledge the risks of participating using the form shown in Section C.1. At the start of each experiment session, participants were led into the motion capture space and asked to wear a motion-tracked glove on their dominant hand. They were asked to stand behind a table, as shown in Figure 6.1, to reduce any likelihood of
injury to participants by restricting their body (except the hand holding the baton) from entering the robot’s reachable workspace.

For each trial, participants picked up the baton off the table and initiated a handover to the robot after hearing the experimenter say ‘go’. Upon detecting the baton in its workspace, the robot would move to retrieve the baton in a way that was consistent with the condition being tested. Three trials were performed for each condition (3 trials * 8 conditions = 24 trials in total per participant). Following each set of three handover trials for a condition, participants were asked to complete the full ROSAS inventory (shown in Section C.2) which asked them to rate how closely each of the 18 items associated with the robotic handovers they just performed. Ratings were on a scale from 1 to 7 where 1 was ‘not at all’, 4 was ‘a moderate amount’, and 7 was ‘very much so’. Each experiment session lasted approximately 30 minutes.

6.5 Results

6.5.1 ROSAS Internal Consistency and Dimensionality

As the ROSAS is a relatively new scale that has not yet been applied to human-robot interactions [25], an internal consistency test is conducted to confirm the results of the exploratory factor analysis performed by Carpinella et al. Internal consistency measures how closely the ROSAS inventory items fit within the three attributes (competence, warmth and discomfort) using the data in this study. For testing, Cronbach’s alpha was used; an \( \alpha_{\text{Cronbach}} \geq 0.80 \) is considered to represent high scale reliability. Items for competence \( \alpha_{\text{Cronbach}} = 0.90 \), warmth \( \alpha_{\text{Cronbach}} = 0.94 \) and discomfort \( \alpha_{\text{Cronbach}} = 0.81 \), all scored above this threshold suggesting that the items have relatively high internal consistency within their respective attributes.

In addition to investigating consistency, dimensionality of the items of each factor of the ROSAS were considered as well. Unidimensionality indicates that the items of each factor measures and corresponds to only one dimension of the scale. On the other hand, if it was found that the items of one factor, e.g., competence, was multi-dimensional (i.e., two or more items are needed to explain a majority of
the variance within that factor), this would invalidate ROSAS as the items do not solely measure competence, but would be measuring something else as well.

A factor analysis was performed to ensure that the items for each attribute are unidimensional. Here, eigenvalues which represent how much variation in each attribute is explained by each item were examined; the larger the eigenvalue, the more variation the item explains. For an attribute to be unidimensional, one would expect to see one item account for a large portion of the variance within the attribute, and other items account for much less variation. As shown in Figure 6.3, the results show that the first items in competence, warmth and discomfort attributes explains 67.7%, 76.9%, and 53.5% of the variance respectively. Given that a majority of the variances are explained by one item within each factor, these findings suggests that the items for each attribute are unidimensional.

**Figure 6.3:** Scree plots for ROSAS factors to examine dimensionality of items.
6.5.2 Effect of Conditions

A three-way repeated measures Multivariate Analysis of Variance (MANOVA) was conducted to test the effect of the manipulated variables (initial arm configuration, speed of retraction, and grasp type) on the ROSAS attributes (Figure 6.4). Effect sizes in terms of partial eta squared ($\eta^2_{\text{partial}}$) are reported; as a rule of thumb, Cohen indicates that partial eta square values of .0099, .0588, and .1379 may serve as benchmarks for small, medium, and large effect sizes [32]. Significant main effects of grasp on reports of competence [$F(1,21)=25.660$, $p<.001$, $\eta^2_{\text{partial}}=.550$] and discomfort [$F(1,21)=7.485$, $p=.012$, $\eta^2_{\text{partial}}=.263$] were found. The latter effect is qualified by a significant interaction effect of speed by grasp on reports of discomfort [$F(1,21)=7.360$, $p=.013$, $\eta^2_{\text{partial}}=.260$]. A post hoc pairwise comparison indicates that the average competence score for the quick [$M=5.225$, $SD=0.980$] grasp type is 1.017 points higher than the mating [$M=4.208$, $SD=1.135$] grasp type [$p < .001$], representing a large effect size [$d=0.835$]. No other main or interaction effects were found to hold statistical significance.

The significant retraction speed by grasp interaction effect (Figure 6.5) was further investigated using paired t-tests at levels of retraction speed ($\alpha = .025$). A significant difference in discomfort scores between the quick [$M=1.648$, $SD=0.638$] and mating [$M=2.580$, $SD=1.140$] grasp types was found at low speed [$t(43)=2.621$, $p<.001$, $d=1.048$]. No significant difference in discomfort scores between quick [$M=2.242$, $SD=1.241$] and mating [$M=2.326$, $SD=1.108$] grasp types was found at high speed [$t(43)=0.370$, $p>.05$, $d=0.072$]. There was also a failure to detect a significant difference between scores at slow and fast retraction speeds for the mating grasp.

6.5.3 Effect of Repeated Interaction over Time

Although the presentation order of conditions was counterbalanced across participants, it was questioned whether participants’ perception changed over the course of repeated handover interactions with the robot. To examine this effect, participants’ trials were categorized by the order in which they were presented in time rather than by experimental condition as shown in Figure 6.6. Trend analysis, a statistical test based upon the F-statistic that is an alternative to an ANOVA [42], was
conducted for each factor with appropriate corrections for non-spherical data. Results showed a significant positive linear trend for warmth \( F(1,21)=7.375, p=.013, \eta^2_{partial}=.260 \) and negative linear trend for discomfort \( F(1,21)=6.442, p=.019, \eta^2_{partial}=.235 \); no significant linear trend was detected for competence. Higher order trends were non-significant for all factors.

**Figure 6.4:** Participants ratings of the robot’s competence, warmth and discomfort over condition as reported during the human-to-robot handover study. Error bars represent 95% CIs.
6.6 Discussion

6.6.1 ROSAS Internal Consistency and Dimensionality

Carpinella et al. claims in [25] that “ROSAS provides a psychometrically validated, standardized measure that can be used to measure robots developed by different people in different places for differing purposes and over time.” As this is the first known usage of the ROSAS for human-robot interaction, it was important to examine the integrity of the scale as it applies to data collected in this study. Although a full validation of the ROSAS using confirmatory factor analysis was not performed due to small sample size, examination of the results show that the 18 items of the scale conform to the three measures of the scale - competence, warmth and discomfort - with a high degree of consistency. Additional testing showed that...
the items of each attribute were highly unidimensional. Thus, the results suggest that the application of ROSAS for this work, and perhaps more generally to other HRIS, appears to be valid - though more work is needed to concretely confirm this.

### 6.6.2 Effect of Conditions

As shown by the results presented in Figure 6.4, grasp type had a significant effect on competence scores, with the *quick* grasp scoring significantly higher than *mating* grasping. This find runs contrary to the expectation that having the robot ensure the handover object’s safety through stable contact would demonstrate more intelligent/competent behaviour. One explanation for this finding is that although the *mating* grasp demonstrates more intelligent algorithms to ensure handover object safety, users may actually find the method to be a significant departure from handovers between human participants compared to the *quick* grasp; thus, they are not able to adapt easily to this novel method of handover. For example, in human-human handovers, receivers apply pulling/tugging forces to the object which signal...
to the giver to release the object [26]. As opposed to the quick grasp, the robot initially applies pushing forces to the object in the mating grasp, which runs contrary to expectation and leads to confusion. As evidence for this, review of video recordings show participants complying to the robot pushing against the baton.

An alternative but complementary explanation for the phenomenon relates to trade-offs made by each grasp type: the quick grasp trades off object safety for efficiency in terms of time to complete the handover, whereas mating does the opposite. Having faster, more seamless handovers may factor more into competence scores than ensuring object safety, particularly if the role of maintaining the object’s safety throughout the handover is the giver’s responsibility rather than the receiver’s as suggested by Chan et al. [26]. In this case, having both participants in the handover be responsible for object safety may feel redundant to the user.

As seen from the results, a significant retraction speed by grasp interaction effect was detected on discomfort scores. Analysis of this interaction effect suggests that for the mating grasp, the discomfort rating was unaffected by retraction speed, whereas the quick grasp increased discomfort to within the same range as the mating grasp in the fast retraction speed condition. This may be due to object safety being doubly compromised by both the quick grasp type and fast retraction which emphasizes speed over safety causing participants to feel that the robot appears too ‘brash’ (as one participant was quoted) in how the object is handled by the robot during the handover. It appears that the quick grasp coupled with slow retraction was rated less discomforting possibly due to increased time for the giver to ensure that the baton is securely grasped by the robot during the retraction phase of the handover. It is possible that discomfort decreased only when both grasp and retraction speed matched their expectations. To further explore if this is indeed the case, analysis of dynamic data was performed and presented in Chapter 7.

As opposed to retraction speed and grasp type, there was a failure to detect any main effects of initial arm position on any of the ROSAS measures. This suggests that user perception of robot competence, warmth and discomfort may be better informed by robot dynamic behaviours rather than static poses. However, it is posited that initial position may still impact on location of the handover, as well as how the negotiation during handover is accomplished. This is explored in the following chapter (Chapter 7).
6.6.3 Effect of Repeated Interaction Over Time

Examination of participants’ evaluations of the robot’s competence, warmth, and discomfort over repeated interactions showed a significant linear increase in warmth and linear decrease in discomfort. Although this may be an effect of repeated application of the scale itself, i.e., participants may tend to centralize their responses in the survey, a linear trend in the force data collected from this study has also been found (presented in Chapter 7). The occurrence of linear trends in both types of data implies that participants are actually changing their perception of the robot, and thus the way they interact with the robot over time. This suggests that the more people interact with the robot, the more they normalize their attitudes towards the robot. The development of familiarity or affinity towards robots is not at all surprising to see as other studies have shown this phenomenon to occur in other contexts such as in assistive home care [62, 76] or military robotics [24]. However, the observation of linear trends in both ratings of warmth and discomfort over time is a notable result. This suggests that changing interaction parameters or attributes of the robot’s receiving gestures could lead to changes in trend rates for warmth/discomfort ratings. If so, these parameters may be tuned or optimized to obtain a fast increase for warmth ratings and decrease for perceived discomfort levels. In turn, this may provide some benefits to having inexperienced users feel comfortable interacting with robots that may appear imposing or foreign - e.g., quickly having factory workers become comfortable working with collaborative industrial robotics. Further study is required.

Failing to detect any significant trends in ratings of competence over time suggests that how competent or able a robot appears to users is not a function of repeated interaction, but rather simply of behaviours attributed to the robot, as seen by the significant main effect of grasp type on competence.

6.6.4 Implications for HRI

Although this work focuses on handovers, some results observed may have wider implications for other HRIS.

- Short of full validation, the ROSAS was shown to be internally consistent and unidimensional across factors. This result is promising in that it suggests that
the scale may be used to similarly evaluate user perceptions of other HRIS, and thus has the potential to serve as a standardized metric for HRI. Additionally, the ROSAS can serve as a valuable tool for designing and evaluating robot appearances and behaviours.

• The effects that were observed with grasp and retraction speed impacting people’s perceptions of the robot’s competence and discomfort may have broader implications for HRI in terms of the trade-offs they present - e.g., faster interaction during HRI may be more efficient, but may cause greater discomfort to users, and seemingly more intelligent behaviours by the robot may not be perceived as such due to decreased efficiency of the interaction. Thus, these observations highlights the importance of considering user perceptions when efforts to develop HRIS which add efficiency or capabilities are undertaken.

• Lastly, the observation of trends over repeated interactions with the robot may not be isolated to just the handover use case. The finding that the more that users interact with the robot, the more they increase their ratings of warmth, while decreasing ratings of discomfort towards the robot may just as likely occur with other HRIS. Thus, this implies that examining inexperienced participants reactions during studies of HRIS may not be as important as considering longitudinal effects and how fast people’s perceptions change over repeated interactions.

6.7 Conclusions

Within this work, factor analysis and examination of the dimensionality of items relating to factors in the ROSAS indicate that the scale appears to be an acceptable tool for evaluating subjective experiences during physical HRI contexts. Thus, in this work, the ROSAS has been used to evaluate user perception of robotic social behaviours during a human-to-robot handover task. Using the ROSAS tool, this work has found that by varying simple parameters such as grasping behaviour and retraction speed of the robot within human-to-robot handover interactions, users can hold significantly different views on social qualities of the
robot in terms of competence and discomfort. Ironically, even though the robot demonstrated a more intelligent grasping strategy in the mating grasp compared to the quick grasp, participants perceived the robot as being less competent and more discomforting. Thus, seemingly intelligent robot behaviours doesn’t necessarily constitute competent or comfortable behaviours in the eyes of users. It appears, rather, that interaction efficiency and/or similarity to human-human handovers (at least in terms of force profiles) constitutes a larger part of establishing more positive user affect when working with the robot. Also, the results of this work indicate that users perceive robots as being less discomforting and having more emotional warmth the more exposure they have to handover over objects to the robot - this may apply to other human-robot interactions as well.

The results presented here offers a glimpse into how users ascribe social attributes to robots during collaborative tasks and how ROSAS can be used to evaluate these perceptions. It is anticipated that the results of this study may inform other human-robot interactions which can be similarly evaluated.

6.7.1 Limitations

Although the sample size of this study was determined to be sufficient for detecting a target range of effect sizes in user perception, a larger sample size could be useful in shoring up the defensibility of findings. Furthermore, the specifics of the experimental setup may limit the generalisability of findings to other human-to-robot handover contexts. In particular, the disparity of user perception between levels of grasp type may be restricted to the use of an electromagnetic end effector and may not be pertinent to other end effector types. A closer examination of other grasping mechanisms and techniques may yield different results.

6.7.2 Future Work

As discussed in the analysis, the results of this study have generated more research questions and numerous pathways for additional examination which can be explored. As only a small subset of non-verbal factors have been explored, future work can be directed towards exploring other factors, e.g., approach speed, laterally-varying initial poses with the robot approaching from the users’ domi-
nant and sub-dominant sides, up/down elbow configuration during handover aim to determine whether similarities or differences exist between human-human and human-robot handovers.

Future studies can also be directed towards determining what differences exist between human-human and human-robot handovers in terms of roles and users’ approaches to handover. Previously, Chan et al. established that both participants in a handover implicitly take up roles during the handover negotiation where the giver is responsible for the safety of the object and the receiver is responsible for the efficiency and pace of the handover [26]. It could be asked whether these roles also exist within the framework of human-to-robot handovers. With regards to the ROSAS, additional work can be done to expand its use for other HRIS, obtain further substantiation of its validity and further explore how social perceptions could/should shape such interactions.
Chapter 7

Exploration of Geometry and Forces Occurring Within Human-to-Robot Handovers

In Chapter 6, a human-to-robot experimental setup was used to examine how changes in the robot’s behaviour (i.e., initial pose, grasp type, and retraction speed) influenced participants’ social perception of the handover via the ROSAS. As a result of that study, it was determined that the way in which a robot exhibits nonverbal cues during handovers has significant impacts on how that robot is perceived.

This chapter continues to explore the design space for human-to-robot handovers by examining how interaction position and force are also affected by the robot’s nonverbal behaviours. Using kinematic and dynamic data gathered from the human-to-robot handover study presented in Chapter 6, handover object geometries (i.e., position and orientation) and dynamics (i.e., forces) can be compared to results obtained from prior work investigating human-to-human handovers - namely work done by Basili et al. [13] and Chan et al. [26]. Additionally, the work in this chapter examines how handover object geometries and dynamics are affected by the nonverbal behaviours of initial pose, grasping method and retraction speed of the robot in terms of interaction fluency and giver/receiver roles as suggested by Chan et al. [26]. Finally, this work studies how repeated interactions with the robot changes force profiles seen during the handover negotiation, draw-
ing ties to similar changes in social perception observed in Chapter 6, suggesting that trustworthiness of the robot may be linked to the forces imparted on the object by human givers.

A manuscript derived from the work in this chapter was presented at the IEEE Haptics Symposium (HAPTICS 2018) held in San Francisco, California, USA from March 25-28, 2018, and is included in the symposium’s proceedings. Some differences exist between this chapter’s contents and the conference paper entitled “Exploration of Geometry and Forces Occurring Within Human-to-Robot Handovers”; these differences are listed below:

- The introduction of this work has been modified to more closely follow the work presented in Chapter 6.
- The background information section has been removed to avoid repeating information found in Chapter 2.
- References to the manuscript examining social perceptions derived from the subjective results of this work have been redirected to Chapter 6.
- As the work in this chapter is derived from the same experiment conducted in Chapter 6, details of the experimental design and setup have been replaced with references to Section 6.3 and Section 6.4 respectively.

Supplementary materials for this work can be found in Appendix C.

7.1 Introduction

In the previous chapter (Chapter 6), a human-to-robot handover user study was conducted examining the effects of changing robot behaviours/nonverbal cues on participants’ social perception of the robot. Specifically, initial arm pose, grasp type, and retraction speed of a robot arm during handover interactions were modified to study how they affected user ratings of the robot’s warmth, competence, and discomfort. The results of the study showed that some robot behaviours, such as grasp type and retraction speed, affected user ratings of competence and discomfort. Additionally, repeated interactions with the robot yielded a bonding/ac-
climatization effect - higher ratings of warmth and lower ratings of discomfort were reported over time.

Here, the same user study is re-examined to investigate physical aspects of the interaction. In particular, this chapter investigates how human givers present the object to the robot (geometry) and what forces they imparted on the object (dynamics) during the human-to-robot handover. Effects of changing the starting pose, grasp type, and retraction speed of the robot are considered as they relate to potentially changing the geometric and kinodynamic negotiation of the object within the handover. Additionally, this work aims to determine if users adapt to repeated interactions with the robot, i.e., to observe learning effects. Through observations of changes in the human giver’s behaviour, the aim of this work is to begin characterizing the design space for human-to-robot handovers from the physical interaction perspective, and be able to inform how subtle alterations of the robot may affect human users.

7.2 Experimental Design

The experimental design and setup used for this study can be found in Section 6.3 and Section 6.4 respectively, and will not be repeated here.

7.3 Geometry

The position and orientation of the baton as held by the participants to initiate handover were recorded with respect to the base coordinate frame of the robot (shown in 6.1). Figure 7.1 shows a scatterplot of positions with X measuring the distance to the robot, Y the lateral offset, and Z the height above ground (not plotted). Figure 7.2 shows the orientation composed of elevation as the pitch angle of the baton relative to horizontal and azimuth as the lateral yaw angle. Rotation along the baton axis was not considered due to symmetry. As the factors of grasp type and retraction speed are non-causal to how participants initially pose the baton (e.g., these factors chronologically occur after the participants’ initiation of the handover and should not affect how the baton is initially positioned and oriented by participants), this data was analyzed with respect to the up versus down initial robot poses.
Only the elevation angle was found to have a significantly differing mean. Participants pointed the baton more horizontally when the robot started in the *up* pose. The *up* pose also led to less variance in both the lateral position and azimuth angle.

![Figure 7.1: Scatterplots of initial object positions (from the overhead perspective) for both the *down* (left) and *up* (right) initial arm pose conditions presented during the human-to-robot handover study.](image)

7.3.1 Results

Paired samples t-tests showed no significant differences in means under *down* versus *up* conditions for all position axes and azimuth angle. Only the mean elevation differed for *down* \([M=14.840°, SD=10.889°]\) and *up* \([M=10.664°, SD=8.209°]\) arm pose conditions \([t(21)=3.470, p=.002, d=0.433]\).

An equivalence test (TOST procedure) was conducted for each position axis using a ±25 mm margin of equivalence. This margin was based on statistical results obtained by Basili et al. in [8] examining the handover object to giver distance for 26 giver/receiver dyads \([M=646.68 \text{ mm}, SD=87.3 \text{ mm}]\). The sample size of 22 was calculated to be sufficient (with a two-sided 90% CIs and 80% power) to establish equivalence, even with a 10% participant loss. Results yielded statistical equivalence between the *up* and *down* groups for all axes at \(p < 0.05\).
Figure 7.2: Scatterplots of initial object azimuth and elevation angles for both the down (left) and up (right) initial arm pose conditions presented during the human-to-robot handover study.

F-tests for comparing variances was performed for each position and angle. The variance in the Y position was significantly different \( F(21,21)=2.149, p=0.043 \), with the variance for the up condition being smaller than for the down condition. Similarly, the variance in azimuth angles differed \( F(21,21)=2.207, p=0.04 \) between down \([M=0.661^\circ, SD=3.621^\circ]\) and up \([M=0.934^\circ, SD=2.437^\circ]\) conditions, again with the up condition leading to a smaller variance. Differences in variance for X and Z positions and elevation angles were not significant.

### 7.3.2 Discussion

**Comparison to Human-Human Handovers**

Table 7.1 compares handover object position results obtained from the human-to-robot experiment with human-human handovers, in particular, as studied by
Basili et al. [13]. Relative to human-to-human handovers, it is observed that participants in the human-to-robot study held the baton approximately 27 cm lower. To explain this discrepancy, it should be considered that the robot apparatus used in this study is approximately 150 cm tall, as seen in Figure 6.1. Meanwhile the average human receiver height appears to be 180 cm in [13], presenting a roughly 30 cm difference. This apparent correlation suggests that givers may be influenced by the proportions of the receiver and place the object conveniently for the receiver. If true, this would imply robots should try to receive handovers at a height proportional to their stature.

### Table 7.1: Comparison of Cartesian positions of handover objects between human-robot and human-human handovers.

<table>
<thead>
<tr>
<th>Direction</th>
<th>Human-to-Robot Mean (mm)</th>
<th>Std. Dev.</th>
<th>Human-to-Human (Basili et al. [13]) Mean (mm)</th>
<th>Std. Dev.</th>
<th>t(46)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>1117</td>
<td>73.36</td>
<td>506.2</td>
<td>131.4</td>
<td>21.446</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Y</td>
<td>-7</td>
<td>54.18</td>
<td>36</td>
<td>53.3</td>
<td>2.764</td>
<td>=.008</td>
</tr>
<tr>
<td>Z</td>
<td>1131</td>
<td>81.988</td>
<td>1407.8</td>
<td>53.7</td>
<td>19.377</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Direct comparisons of lateral and distal positioning within the horizontal plane is challenged by differences in experimental procedure. For example, the protocol of the study presented here includes a table that was placed between giver and robot receiver which was not used in Basili et al.’s experiment. Additionally, right- as well as left-handed handovers were allowed, whereas only right-handed handovers were permitted by Basili et al. Nevertheless, regardless of whether the receiver is human or robot, both studies observe that the handovers occur roughly halfway and centered between giver and receiver.

**Effect of Initial Robot Pose on Geometry**

The results of the study shows that the initial robot pose significantly affects a giver’s placement and orientation of an object for handover. When the robot starts
in the up pose, thus closer to giver and the eventual handover location, a giver more tightly places and orients the object in the horizontal plane. They also lower the elevation angle, more aligned with though still significantly above the robot’s end-effector angle of $5.94^\circ$ [$t(21)=2.698, p=0.014, d=0.575$]. It appears that givers generally attempt to place and orient the object complimentary to the end-effector, at least as much as is comfortable.

This observation agrees with arguments made by Cakmak et al. maintaining that the spatial configuration may be an important tool for improving handover interaction fluency through implicit, nonverbal communication [22]. Thus, this result suggest the robot’s up pose implicitly communicates to users, better informing them where and in what orientation the robot can reach for the object. Such communication is particularly important in human-to-robot interactions. Where human givers likely have lots of experience handing objects to human receivers, they may have limited a-priori understanding of robot handovers. Especially in the down pose, the amorphous shape of the KUKA LBR iiwa provides few cues and givers may remain uncertain how to present the object. Possibly poor placements could then require longer robot trajectories or awkward grasp angles, limiting fluency and efficiency of the interaction. Thus, it is hypothesized that the robot’s up pose, illustrating the preferred handover angles and location, increases handover fluency and efficiency.

7.4 Dynamics

Interaction forces during the handover was captured using the force/torque sensor attached to the robot’s end-effector. For the purposes of this study, however, only the forces applied axially with respect to the end-effector were analyzed. Inertial and gravitational components were subtracted from the data using the observed kinematics to calculate the isolated interaction forces experienced by the human giver. Additionally, data was filtered using a fourth-order low-pass Butterworth filter with a 14 Hz cutoff, similar to [26].

For this analysis, the maximum retraction/pull force applied by the robot to the giver was considered. It has been postulated that this absolute level communicates that the receiver is in full control of the object and triggers the giver’s release.
The maximal change in retraction force, relative to the force immediately before retraction is also considered as relative changes may provide additional information and triggers to the giver. These metrics are illustrated in Figure 7.3. Pulling forces applied to the end effector are denoted as positive, whereas pushing forces are negative.

**Figure 7.3:** Sample end effector axial force vs. time plot for a participant depicting features used in the dynamics analysis of the human-to-robot handover study. Negative forces indicate pushing (compressive) forces exerted against the end effector whereas positive forces indicate pulling (tension) forces.

### 7.4.1 Results

The mean maximal absolute and relative retraction forces are depicted in Figure 7.4. In particular, the overall mean maximal absolute retraction force was 5.48 N \(\pm 7.11\text{N}\) or 223% \(\pm 290\%) of the baton’s 250 g weight. A three-way repeated measures MANOVA was conducted to test the effect of the manipulated variables (initial arm configuration, speed of retraction, and grasp type) on the mean maximal absolute and relative retraction forces. Effect sizes in terms
of partial eta squared ($\eta_{\text{partial}}^2$) are reported. Results showed significant main effects of grasp type $[F(1,21)=9.765, \ p=.005, \ \eta_{\text{partial}}^2 = .317]$ and retraction speed $[F(1,21)=10.322, \ p=.004, \ \eta_{\text{partial}}^2 = .330]$ on mean maximal absolute retraction forces. For the relative retraction forces, a significant main effect was only observed for retraction speed $[F(1,21)=10.888, \ p=.003, \ \eta_{\text{partial}}^2 = .341]$. No other main or interaction effects were found to be significant.

Figure 7.4: Maximum absolute and relative retraction forces as experienced by the giver in the human-to-robot handover study. Asterisks represent significant comparisons at the $p < 0.05$ level and error bars represent 95% CIs.

### 7.4.2 Discussion

#### Comparison to Human-Human Handovers

Chan et al. report that human givers tend to delay the release of an object even after the receiver is fully supporting the object’s weight [26]. They measured a maximum excess receiver load and hence a positive maximum retraction force of 2.36% [$SD=4.16\%$] of their baton’s 483-678 g weight. With the receivers pulling
more than the object’s weight, they hypothesized this may be a precautionary behaviour on the part of the giver to ensure safe object transfer.

For a robotic receiver, a nearly 100-fold increase in this metric is observed. Following the above hypothesis that a giver only releases the object when they believe safety is guaranteed, this could imply participants were not as confident or trusting in the robot receiver. Such a lack of confidence would be consistent with inexperience in human-to-robot handovers. But this argument would necessitate that the interaction forces are only created by voluntary giver actions.

An alternative explanation would come from involuntary forces. If the robotic retraction follows a different timing, motion profile, speed, or even impedance than a human retraction, the interaction forces might also differ without any voluntary consideration. For example, the retraction forces could be generated before the receiver has a chance to react. As such, this could suggest an efficient human-to-robot handover will require subtle retraction movements.

Effect of Grasp Type

The grasp type had a significant effect on the maximal absolute retraction force, with mating grasps resulting in approximate half the force of quick grasps. Following the above logic, this could signify that participants felt more trusting of the mating grasp and thus released the object at a lower absolute force threshold.

However, recall that during the mating grasp, the robot initially applies a pushing force in an attempt to obtain flush contact. Meanwhile in the quick grasp, the magnet is already pulling the object. Indeed, the maximal relative retraction force does not show a significant difference between the two conditions. This could suggest that givers are relatively indifferent to the grasping type and trigger their release on a relative force change. And as before, any involuntary reaction forces may compound the observations.

Effect of Retraction Speed

The slow and fast retraction speeds may shed the most light on the issues of involuntary force buildup. Both the maximal absolute and relative retraction forces were significantly affected by the retraction speed condition. In particular, twice
the retraction speed resulted in nearly twice the retraction force. Also several participants noted that in the fast condition, they felt the robot yanking the baton out of their hand.

The differences in force profiles may have less to do with voluntary force thresholds and more with human reaction time, which is on the order of 150 ms for haptic stimuli. In the slow condition, the maximal retraction force occurs 147 ms \([SD=50\text{ ms}]\) after the start of the retraction. In the fast condition, the timing is much shorter. A fast retraction thus exceeds the giver’s grip forces before they can react. To avoid any sensation of yanking and generally to allow the giver to voluntarily control forces, releasing the object as appropriate, the robot receiver will need to carefully modulate and limit retraction speeds. At least until the robot can determine that the object has been released.

Finally, it is noted that the apparent correlation between forces and retraction speeds suggests that the givers are presenting repeatable impedances during handover. Such findings could also help guide robot behaviours in handover to a human.

7.5 Learning

Although the presentation order of conditions was counterbalanced across participants, each participant’s perception was expected to change over the course of their repeated handover interactions with the robot. To examine this effect, participants’ trials are shown in chronological order in Figure 7.5.
Figure 7.5: Maximal forces during handover as experienced or applied by givers during the human-to-robot handover study over trials ordered chronologically. Lines of best fit for each are shown. Error bars represent 95% CIs.
7.5.1 Results

Trend analysis was conducted for each factor with appropriate corrections for non-spherical data. Results showed significant negative linear trends for maximal absolute \( [F(1,21)=11.924, p=.002, \eta^{2}_{partial} =.362] \) and relative \( [F(1,21)=7.607, p=.012, \eta^{2}_{partial} =.266] \) retraction force. Higher order trends were non-significant for all measures.

7.5.2 Discussion

The observation of negative linear trends in both force measures over repeated interaction with the robot is notable as it indicates that participants are adapting their force behaviour to the robot. If voluntary behaviour is considered, the givers may be building up trust in the robot to safety receive the object and releasing sooner. Alternatively, if involuntary forces are considered, givers may be learning to predict the robot’s behaviours and moving or relaxing predictively without necessarily releasing sooner. Indeed these two aspects may be fundamentally linked in the human givers, as the ability to predict would seem to go hand in hand with any willingness to trust.

Further evidence of learning can be derived from participant’s subjective ratings of the robot during the experiment through the ROSAS inventory: ratings of the robot’s warmth linearly increased over repeated interactions, while discomfort simultaneously decreased (see Sections 6.5.3 and 6.6.3). This suggests that the more people interact with the robot, the more they develop positive attitudes towards the robot. Both warmth and discomfort are known factors in the determination of trustworthiness of both humans and robots \([33, 82]\). Thus, when considering both sets of trends together - decreasing force and increasing positive social perception of the robot - there appears to be strong evidence that forces imparted on the object by the giver are related to how willing they are to trust the robot with the safety of the object. Although, again, it is unclear whether lower forces cause higher ratings of warmth and decreased discomfort (or vice versa), or whether both are effects of another factor at play, e.g., of familiarity or predictability of the robot.
7.6 Conclusions

Beyond providing a demonstration of a simple human-to-robot handover, the user study was able to elicit some basic lessons on appropriate behaviour. First, the results of the study showed that the robot’s initial pose affects the handover geometry. As a result, it is posited that a pose can communicate appropriate location and orientations for the handover, information that may not be obvious to an inexperienced giver. As such, the initial pose influences interaction fluency. Additionally, evidence was found to indicate that givers may cue off the robot’s height, as they would off a human receiver’s height.

The examination of interaction forces suggests that givers release the object when they detect an appropriately large change in retraction force. That is, an increase in the force by which the robot is pulling would indicate it is securely holding the object and trigger the release. One theory is that this level depends on the giver’s general trust in the robot’s ability to grasp the object; however, as there is no direct measure of this trust, direct correspondence cannot be established at this time.

Results also show a main effect of retraction speed which caused significantly larger interaction forces for faster retraction speeds. It is posited that a fast retraction preempts the giver’s ability to react to the withdrawal. The robot simply overcomes the grip forces and yanks the object away. Thus, to provide a refined handover experience, human reaction time must be considered and retractions must be modulated carefully.

Compared with human-to-human handovers, the interaction forces were generally significantly higher. This difference could be attributed to the inexperience of participants in handing over to the robot; participants may seek a larger force to confirm that the robot has securely received the object. Symmetrically, this might suggest that the robot is not acting exactly like a human receiver and hence presenting unexpected or unpredicted movements. Over time and repeated interactions, however, this effect and the force levels linearly decrease. Separate social perception evaluations (using the ROSAS) mirror this trend with a significant linear increase in warmth and linear decrease in discomfort. Together this may indicate
givers are learning to predict the robot and developing trust in the robot to complete the handover successfully.

These findings show that slight changes to robot behaviours may significantly alter interaction dynamics of the negotiation that occurs during handovers: we have found significant differences to the way users kinodynamically participated as givers during the handover through varying three robot attributes. The results of this work more generally suggests that an exploration of the design space for human-to-robot handovers may assist in achieving more fluent and legible, though not necessarily human-like, handovers. Such improvements in handovers (and ostensibly other human-robot interactions) may be measurable through examination of interaction geometries and forces, as demonstrated here.

7.6.1 Future Work

As discussed in the analysis, the results of this study have generated more avenues of work to be explored. Specifically, future studies can more definitively address how repeated handovers with the robot affect force levels and, more broadly, trust in the robot with respect to the object’s safety. Another related research topic might be an investigation of how perceived value and/or fragility of the object might impact the necessary force levels imparted on the object by the user during a human-to-robot handover. Also, what factors can be changed or improved regarding the robot’s behaviour that might allow users to trust the robot more quickly? Future studies may also address how robot height and appearance, as well as contact impedance and movement fluidity may impact the interaction.
Chapter 8

Conclusions

Being able to achieve seamless object handovers between robots and humans offers many opportunities for closer human-robot teaming and collaboration. For example, in manufacturing, the handover of tools or parts between robot-worker pairs can improve productivity and cost-savings. Alternatively, in assistive care for the elderly and infirmed, robotic companions could help retrieve and handover hard to reach objects such as a bottle of pills or television remote, enhancing the autonomy of patients. In these and other cases where physical cooperation between a robot and human can augment the person’s abilities and efficiency, having a robot be able to pass or receive an object to a person can be an especially useful ability.

Towards the goal of enabling robotic agents to participate seamlessly in handover interactions, this thesis explored how robots might both recognize and display nonverbal cues to facilitate object handovers to and from a human. A series of six human-subject studies collectively explored the following two research questions:

Q1 How do nonverbal cues exhibited during robot giving and receiving behaviours change how users perceive the robot, and affect the handover negotiation?

Q2 How can a robot adequately recognize and interpret nonverbal cues conveyed by a human to infer object attributes as well as handover intent?
Results from the six studies support that nonverbal cues offer effective means to improve human-robot handover interactions in terms of fluency, efficiency, and user perception. Additionally, in the case of human-to-robot handovers, we also find that detecting nonverbal cues from a human giver can enable a robot to recognize handover intent. The remainder of this chapter presents a summary of the findings, implications, and future work that pertain to each of the questions.

8.1 Conveying Non-Verbal Cues During Handover

8.1.1 Summary of Findings

Non-verbal cues of a robot have been used in various HRI contexts to establish joint attention with and communicate a robot’s intent and internal states to a user. A large portion of this thesis focused on the role nonverbal robot behaviours can play during handovers. A summary of the cues that have been studied can be found below in Table 8.1.

Gaze

The investigations presented in Chapter 3 examined how gaze cues can signal handover intent, location and timing, thus establishing fluency and legibility of handover interactions. From a human-human handover study, givers were found to use a small set of gaze profiles. Of particular interest were two gaze profiles named the Shared-Attention and Turn-Taking gazes. In the Shared Attention gaze profile, the giver would direct his/her gaze towards the future handover location as he/she was moving the object to that location. In the Turn-Taking gaze, the giver looks up and directs his/her gaze towards the receiver’s face following a Shared Attention gaze and completion of trajectory placing the object at the handover location. From observation of these two gaze profiles being used, it is posited that they serve distinct purposes in increasing fluency of the handover interaction - the Shared-Attention gaze serves to inform the receiver where the handover is to take place, whereas the Turn-Taking gaze informs when it is appropriate for the receiver to take the object.
Table 8.1: Summary table of robotic nonverbal cues studied in this thesis and their observed impacts.

<table>
<thead>
<tr>
<th>Non-Verbal Cue</th>
<th>Handover Type</th>
<th>Impact(s) on Handover</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze</td>
<td>R→H</td>
<td>Handover timing (Efficiency)</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>Initial Arm Pose</td>
<td>H→R</td>
<td>Giver’s placement and orientation of an object for handover (Fluency and Efficiency)</td>
<td>Chapter 7</td>
</tr>
<tr>
<td>Grasp Strategy</td>
<td>H→R</td>
<td>Perceived competence and discomfort of robot (Fluency)</td>
<td>Chapter 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maximum retraction force imparted on handover object</td>
<td>Chapter 7</td>
</tr>
<tr>
<td>Object Retraction Speed</td>
<td>H→R</td>
<td>Perceived discomfort of robot (Fluency)</td>
<td>Chapter 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maximum retraction force imparted on handover object</td>
<td>Chapter 7</td>
</tr>
</tbody>
</table>

A second study was conducted which had a robot giver imitate both the Shared Attention and Turn-Taking gaze profiles, as well as a control gaze profile where the robot gazed downwards for the entire interaction (No Gaze). This study explicitly addressed the impact a robot’s gaze has on human behaviours during a handover, measuring subject perception and reaction time of human receivers. Results of these robot-to-human handovers demonstrated that the robot’s use of gaze can impact when human receivers decide to move their hands to receive the object from the robot. Participants were observed to reach out for the object proffered by the robot significantly earlier when the robot exhibited the Shared Attention gaze than the No Gaze condition which did not use a human-inspired gaze pattern. This suggests that the implementation of nonverbal cues on a robot, such as gaze, can influence people’s behavioural responses to an interaction while the interaction is taking place. In addition, with the Shared Attention gaze, participants reached to the projected handover location before the robot had fully reached the location. By
cueing the participants to reach out earlier to meet the proffered object before the robot has finished moving, the interaction is not only more efficient but also more fluid.

**Exploring Non-Verbal Cues in Human-to-Robot Handovers**

Chapters 6 and 7 explore how robot nonverbal cues during human-to-robot handovers may affect user perception of the robot, geometry (i.e., position and orientation) of the handover object, and negotiation forces experienced during the interaction. Unlike the other studies mentioned in this thesis, this work does not explicitly draw upon inspiration from human-human interactions. The study presented in both chapters features a disembodied robot arm which received a handover baton from human participants during handovers. Initial arm pose prior, grasping behaviour during, and arm retraction speed following handover were varied in a 2x2x2 factor experimental design.

Through survey of participants’ perception of the robot across different conditions using the ROSAS, results showed that participants perception of the robot’s competence and discomfort were significantly affected by grasp type. It was found that a ‘brute force’ approach to grasping, where the robot energizes its electromagnetic gripper when it comes to within 1 cm of the handover baton (quick grasp), was rated as more competent and less discomforting (in the case of slower retraction speed) compared to a grasping behaviour which carefully, yet more slowly, obtains flush contact by pushing against the baton before activating the magnet (mating grasp). One possible explanation for this result is that although the mating grasp demonstrates more intelligent behaviour to ensure handover object safety, users may actually find the method to be a significant departure (in terms of force profile) from human-human handovers compared to the quick grasp. Thus, participants may be associating expectations of how handovers should occur to competency, and since they are not able to adapt easily to this difference in expectation in the case of the mating grasp, they rate the interaction as less competent. Alternatively, participants may also be associating interaction efficiency with competency, where the simpler, quick grasp type outperforms the more complicated mating grasp type because of its speed and no-frills grasping algorithm. A take
away of this result is that robot behaviours deemed intelligent by the designer do not necessarily constitute competent or comfortable behaviours from the perspective of users. Thus, users should be consulted on the design of robot behaviours.

In Chapter 7 results showed that initial pose affects the handover geometry. Specifically, participants tended to reduce variation in position and orientation of the handover object (baton) when the robot was posed in a way that conveys the robot is awaiting the handover object (up pose) compared to when the robot arm was positioned hanging straight down (down pose). Thus, it appears that initial pose of the robot can intrinsically communicate appropriate geometries for the handover, leading to improved interaction fluency and efficiency.

When examining interaction forces occurring during handovers, it appears that givers release the object when they detect a large change in retraction force - i.e., a relative threshold. In this case, an increase in the force imparted on the object as the robot is retracting it indicates to the giver that the object is being securely held, and can let go without fear of the object dropping. Although it is hypothesized that this threshold depends on trustworthy the human giver feels the robot is with respect to object handover, testing this theory is left as future work. When comparing the magnitude of forces experienced in human-to-human against human-to-robot handovers, the forces were found to be significantly higher in the later case - on the order of a 100-fold increase. The inexperience of participants in handing over to the robot could explain much of this difference as participants may use a larger force threshold to confirm that the robot has securely received the object before releasing it - having the robot behave unexpectedly compared to what human givers may be familiar with in human-to-human handovers may cause confusion as to how the giver is to behave in terms of the force negotiation involved in handover over the object to the robot. Lastly, results also show that faster retraction speeds caused significantly larger interaction forces. From examination of force data, it appears that fast retraction of the object causes the robot to simply overcome the grip forces imparted by the user and yanks the object away before the user can react. According to participant subjective reports, this leads to a more discomforting interaction. Thus, this suggests that to provide a refined handover experience, human reaction time must be considered and post-grasp retractions must be modulated carefully.
8.1.2 Implications

These results from these studies supplement the previous work in examining human-human and human-robot handovers discussed in Chapter 2. In previous studies, nonverbal cues were shown to be effective in communicating a robot’s target object and internal states to an observing person. Findings from this thesis add to prior work in that they demonstrate that nonverbal cues such as gaze and object orientation in the case of robot-to-human handovers, and initial pose, grasp type, and retraction speed during human-to-robot handovers, can significantly affect multiple aspects of the interaction; such aspects include user perception, fluency, legibility, efficiency, geometry, kinodynamics and fluidity of the handover. Thus, the work presented in this thesis indicates that nonverbal cues can serve as a powerful medium by which details of a handover can be subtly communicated to the human receiver and elicit desired kinodynamic behaviours.

The implications of this work suggest that those who design behaviours for HRI, must be cognizant of how nonverbal cues can improve interactions between collaborating robots and humans. Non-consideration or implementation of poorly designed cues may be significantly detrimental to interaction efficiency, fluency, fluidity and legibility; whereas considering where and how cues can be used, perhaps through human-inspired means, can drastically improve how people perceive and interact with the robot. Results of studies presented in this thesis also indicate that the design space for nonverbal cues within handovers is potentially quite vast and largely unexplored. Although much of the prior literature uses human nonverbal behaviours as a framework for the design and exploration of cues within this space, deviating from this approach (much like how robot grasp type was investigated in Chapters 6 and 7) may allow researchers to investigate unique and unexpected robot cues yielding alternative interaction cues for human-robot handovers and other interactions.

8.1.3 Limitations

Various nonverbal cues may signal differently to people of various ethnic origins and backgrounds. Thus, one caveat of this work is that the results may only be applicable to populations within North America. Much more extensive studies
involving various contexts and nonverbal cues using other populations will be necessary before being able to draw generalized conclusions about the roles a robot’s nonverbal cues can have on characteristics of human-robot handovers.

These studies mainly relied upon non-expert users having little to no experience or background with robotics. Thus, the reported results may not be applicable to expert users having an understanding of the limitations and peculiarities of the robots they are interacting with. Backed by some findings reported in this thesis that a learning effect has been detected (particularly for work presented in Chapters 3, 6 and 7), it would be expected that the obtained results would differ in the case where employing expert participants were used for the studies.

As a related limitation, all of these studies employed physical barriers (e.g., table, low wall) to limit non-expert users from approaching too closely to the robots in an effort to curb unintentional and potentially injurious physical encounters. While being able to limit users’ physical exposure to the robots and also satisfying safety concerns of behavioural research ethics review boards, the resulting experimental setups are sub-optimal for exploring human-robot interactions that might realistically occur in the field, where humans and collaborating robots are working in close proximity. As such, future experiments in this and other HRI contexts could be aimed towards having expert users work with robots without needing safety barriers.

While limited in scope, however, the results of the work presented in this thesis provides empirical support that even the subtle, and supplementary nonverbal cues displayed by a robot can play a significant role in object handovers.

### 8.1.4 Future Work

In order to further explore how nonverbal cues may affect handover interactions, a method of study which may provide interesting results would be to intentionally misdirect or mislead participants using nonverbal cues. For example, in terms of the gaze study found in Chapter 3, if the robot intentionally gazed upon a location other than the handover location, would that adversely affect the duration of the handover? Additionally, with reference to the dynamic interactions during human-to-robot handovers reported in Chapter 7 another study could be conducted
which examined if human givers actually employ a force thresholding technique in
determining when to let go of the object. In such a study, the electromagnetic gripper
would randomly fail to activate prior to the robot arm retracting - if participants
are indeed using force thresholding, the giver would continue to hold onto the ob-
ject following retraction, whereas if another methodology is employed, the giver
may drop the object altogether.

As a result of observing linear trends in user perceptions of competence and
warmth (Section 6.6.3) and interaction forces (Section 7.5) over repeated interac-
tions with a robot, another avenue of future work may address how repeated han-
dovers with the robot affects forces experienced during the negotiation, and more
broadly, trust in the robot with respect to the object’s safety. Specifically, what fac-
tors can be changed/improved regarding the robot’s behaviour that can have users
become more trusting of the robot more quickly?

8.2 Recognizing Non-Verbal Cues from Humans

The second theme and research question of this thesis dealt with having robots
recognize and interpret nonverbal cues from humans. In relation to this theme,
one chapter (Chapter 4) examines the quality of nonverbal cue content naturally
generated by humans (object orientations during handover in particular) as it per-
tains to using such observations of human behaviours for training robot behaviours.
Another chapter (Chapter 5) investigates how robots can infer handover intent from
human kinematic behaviours to enable human-to-robot handovers.

8.2.1 Summary of Findings

Object Orientation

Chapter 4 presents a human-human study where the orientation in which house-
hold objects are presented by human givers to human receivers are examined. The
goal of this work was to determine if human-human handover data could be used to
teach robots how to appropriately hand over objects. Givers were asked to handover
twenty household objects under three conditions: without instruction (natural), in a
fashion that was considerate of him or herself as the giver (giver-centered), and in a
fashion that was considerate of the receiver (receiver-centered). Mean orientations of the objects as used by givers were calculated using an optimization function for the three types of handover conditions. Results of this study showed patterns in the way participants oriented objects during handovers where objects were aligned by a specific axis associated with the object across multiple participant pairs and handover trials. This pattern was particularly prominent in the receiver-centered handover condition. Postulating that these patterns arise from the affordances of the objects, the concept of an affordance axis as well as a mathematical definition was developed to track these alignments across conditions. From comparison of the affordance axes between receiver-centered and giver-centered handovers, significant differences in orientation were detected. This finding suggests that mean orientations and/or affordance axes of objects may be useful in teaching robots how to appropriately align objects during robot-to-human handovers.

Natural handover object orientations were found to be significantly different from orientations observed in the other conditions for a majority of the objects tested, perhaps due to mixing of receiver- and giver-centered handovers. This implies that a robot will need to consider the quality of naturally-observed handover orientations when learning from them. Variation of orientations may provide a metric for quality could be as objects that have different natural and receiver-centered handover object orientations appear to exhibit larger variance during natural handovers.

**Recognizing Handover Intent**

The question of whether nonverbal cues of a human giver be used to infer handover intent by a robot is addressed in Chapter 5. In developing a system for recognizing handover intent, machine learning classifiers were used on observation of kinematic behaviours due to the widespread availability of sensors that could track human motion and kinematics (i.e., the Microsoft Kinect). A human-human handover study was conducted where kinematic behaviours of the giver were recorded via a motion capture setup. Overall, 176 features were obtained from the motion capture data, however, a systematic, bagged random forest approach was used on a training set to select a smaller set of 22 features deemed to be highly discrimi-
native between handover and non-handover motions. These 22 features were then used to train several SVM models with different kernels to detect handover intent. Test results on a holdout partition indicated an $>80\%$ accuracy for all kernels, and a maximum accuracy of 97.5% by the SVM with an RBF kernel in its capacity to detect handover motions.

To further demonstrate feasibility of this method using more commonly found motion tracking sensors employed by robots, an SVM was similarly trained on a Kinect version 2 and tested in a pilot study. The results of this study showed that handover detection performance was similar to that seen with the models trained using the motion capture data.

### 8.2.2 Implications

When taken altogether, these results demonstrate considerable potential for the detection of object properties, handover events and other gestures for HRI using kinematic features, whereas prior methods used include image analysis and manual coding of time-series data. The ability for a robot helper to accurately recognize elements of handovers provides a solid, first step in allowing robots to appropriately determine how to handover objects efficiently, react to human handover gestures, and receive objects from people. Also, the findings of these studies build on top of prior work highlighting the feasibility of recognizing human nonverbal cues which enable autonomous behaviours within the context an HRI task. Having robots recognize what a collaborating human agent is trying to nonverbally communicate, and determining how best to assist this person provides a powerful capability which can allow robots to be much more valuable in a wide variety of environments.

### 8.2.3 Limitations

Experiments and data collection for these works was performed in a controlled laboratory setting where the handover procedure was structured. Thus, the results obtained may not be representative of what might occur in the field. Sample sizes of these studies are relatively modest, which may limit the generalisability of these findings.
8.2.4 Future work

Given that several types of nonverbal cues are recognizable by robots, one course of future work would be to have robots learn, from human demonstration, to trajectory handover giving and receiving motions such that they may appear fluent and possibly human-like to human collaborators. One method which might enable this is the use of Dynamic Movement Primatives (DMPs) to both learn motions and adapt to variation in handover parameters (i.e., handover location) in real-time \[102\]. A similar treatment might also be given to how a robot renders forces on the shared object as it is being passed from giver to receiver in a handover negotiation - such an effort may borrow from prior work in haptics demonstrating observation and mimicry of force profiles \[69\].
Bibliography


doi:10.1007/978-3-540-35488-8. → pages 88


doi:10.1109/TIM.2011.2161140. → pages 80, 107


doi:10.1023/B:JINT.0000021039.56110.c8. → pages 31


172


Appendix A

Supporting Materials for Investigating Robot Gaze in Robot-to-Human Handovers

A.1 Study Advertisements
A.1.1 Poster

COME GET A DRINK
served by our ROBOT

It’s a Human-Robot Interaction Study!

• Get a Drink!
• 5 Minutes!
• Help Science!

our robot will hand you some objects,
you will tell what you think of it
A.1.2 E-Mail

Name: UBC PR2 Hackathon Team, Collaborative Advanced Robotics and Intelligent Systems Laboratory
Department: Mechanical Engineering and Computer Science
E-mail Address: [Redacted]
Event Name: Get a drink from our robot
Event Date: September 3, 2013 (Imagine Day)
Time: 10am 2pm
Schedule of Events: Drop by any time between 10am 2pm
Location (Include building name & room number): ICICS building, Room 146
Event Description: Thirsty? Want to see a robot? Come find out what it is like to get a drink (bottled non-alcoholic beverage) from a robot. We are conducting a very short (5 minute) experiment to study how robots should handover objects to people. All you need to do is to drop by to get a free drink from our robot and fill out our questionnaire to tell us what you think about it. We will not pay you for participating in this study.

We need lots of people to come participate. So bring your friends! Any University of British Columbia (UBC) student of 17 years of age or older, and anyone (even non-students) of 19 years of age or older can participate.
Registration Required (Yes/No): No
Who to Contact for More Information: AJung Moon [Redacted]
This event is open to: All UBC students over 17 years of age, and all faculty, staff and visitors of 19 years of age or older.
A.2 Consent

A.2.1 Written Consent

HRI-Cues: Human-Robot Handover Study Consent Form

Project Title: HRI-Cues: Human-Robot Handover Study
Principal Investigator: Dr. Elizabeth Croft, [REDACTED]
Contact: AJung Moon, [REDACTED]
Funding: This research is funded by the National Sciences and Engineering Research Council of Canada (NSERC).
Purpose: The purpose of this project is to investigate how robots should hand over objects to people. Results from this study will help us develop robots that can interact with people better.
Procedures: The entire experiment will take no longer than five minutes. In order to participate in this study, you must be 19 or older, or a UBC student of 17 or older. You will be asked to stand at a designated spot in front of a robot. The robot will hand over one or more bottled non-alcoholic beverages for you to take. After each time the robot gives you a beverage, as well as at the end of the experiment, you may be asked to fill out a questionnaire about the handover. You may keep one of the beverages the robot will give you, but we will not pay you
for participating in this study. You may refuse to participate in this study and you may withdraw at any time by exiting the room without interacting with the robot. 

**Potential Risks:** You may physically come into contact with the robot. This robot has been designed for safe human interaction and robot speeds and forces will be kept to safe levels. The handover has been tested and found to be safe in earlier studies.

**Confidentiality:** No identifying information will be collected or stored with your data, in order to ensure your privacy. This study will be video recorded, including your face, for analysis purposes. Video recordings from the experiment may be presented at scientific conferences or published in reports/journals, but identifying features (including your face) will be blurred using digital blurring tools. Only the researchers involved with this study will be able to view the unblurred video recordings. Data collected during the experiment will be stored on a password protected computer in the CARIS Lab, which has restricted secure access and is locked at all times.

If you have any concerns about your treatment or rights as a research subject, you may telephone the Research Subject Information Line in the UBC Office of Research Services at the University of British Columbia, at (604) 822-8598.
A.2.2 Verbal Consent

Consent script to be read by co-investigators and/or additional study team members.

This is an experiment being conducted by the CARIS lab to investigate human-robot handovers. You will be asked to stand in a designated location and a robot will hand you one or more bottled non-alcoholic beverages. After each handover, as well as at the end of the experiment, you might be asked to answer a few short questions. The study will be video recorded, but your face will be blurred before the video is shown outside of our research group.

You may withdraw from the study at any time by exiting the room.
You may keep one of the beverages the robot gives you, but you will not be paid.

Here is detailed information about the study, including contact information if you have any questions or concerns.

[Hand participant the printed consent form (refer to Section A.2.1)]

Do you agree to participate in this study?

[Obtain verbal consent]

Are you 19 or older, or a UBC student of age 17 or 18?

[Obtain verbal agreement]

Do you agree to be video recorded for this study?

[Obtain verbal consent]

Do you have any questions?

[Obtain any questions]
A.3 Surveys

A.3.1 Single Condition Survey

Subject Number: ______
Handover: ______

What is your age? ______
What is your gender (circle one)? M  F

For each question, circle the number that represents how you felt about the handovers.

1. I liked this handover.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

2. The handover seemed natural.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

We think that timing is important in human-robot handovers. Considering the timing in the handover, please answer the following:

3. It was easy to tell **when**, exactly, the robot wanted me to take the object.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

Any additional comments?

__________________________________________________________________________

__________________________________________________________________________

Last Revised: August 7, 2013

Rev 1
A.3.2 Condition Comparison Survey

Subject Number: _____
Handover 1: _____
Handover 2: _____

What is your age? _____
What is your gender (circle one)? M   F

1. Which handover did you like better, overall (circle one)?

   First Handover   Second Handover

2. Which handover seemed more natural (circle one)?

   First Handover   Second Handover

3. We think that timing is important in human-robot handovers. Which handover made it easier to tell when, exactly, the robot wanted you to take the object (circle one)?

   First Handover   Second Handover

Any additional comments?
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Appendix B

Supporting Materials for Investigation of Handover Object Orientations and Automated Detection of Handovers

B.1 Study Advertisements
Volunteers Wanted for Human-Robot Interaction Study

Project Title: HRI-Cues: Human-Human Handover Study II
Principal Investigator: Elizabeth Croft
Co-Investigators and Contact Persons: Matthew Pan, Wesley Chan

The CARIS Laboratory in the UBC Department of Mechanical Engineering is seeking healthy adult volunteers to participate in a study investigating how humans hand objects to each other. The knowledge gained from this study will be used in our ongoing work to develop intelligent robotic assistants.

The study will consist of two participants handing objects back and forth. We encourage participants to sign up in pairs, but all volunteers are welcome. Please include “HRI-subject” somewhere in the subject line of your email, and thank you for your interest.

This study will run throughout November 2014. For information regarding this study or to volunteer as a participant, please contact:

Matthew Pan
ICICS X015, 2366 Main Mall
mpan9@interchange.ubc.ca

Please place “HRI” somewhere in the subject line. Thank you for your interest in this work.

Matthew Pan, PhD Student, Mechanical Engineering, UBC, Wesley Chan, PhD Student, Information Science & Technology, University of Tokyo, Elizabeth Croft, Professor, Mechanical Engineering, UBC,
B.1.2 E-Mail

Re: Call for Volunteers for a Human-Robot Interaction Study

The Collaborative Advanced Robotics and Intelligent Systems (CARIS) Laboratory is conducting a study in human-robot interaction. We are seeking healthy adult volunteers to help us better understand how humans hand objects to each other. Our aim in this study is to characterize this behaviour and identify cues that humans use to coordinate object transfer. The knowledge gained from this study will be used to develop robots capable of handing objects to humans and vice versa safely and intuitively.

The study will be conducted in the Human Measurement Laboratory (room X527) in the ICICS building, and will require approximately 1 hour of your time. Participants will be paired up and asked to hand a several household objects and forth while wearing motion capture markers. You will be asked to repeat the handover task with different objects each time. Non-identifying motion capture data will be collected. The experiment will also be videotaped if both participants consent. All collected data will remain anonymous. You may sign up for this study with another person as a pair or by yourself.

The study will take place throughout the month of November. For more information regarding this study or to volunteer as a participant, please contact:

Matthew Pan, or
Wesley Chan,

Thank you for your interest.

Matthew Pan, Ph.D. Student, UBC Mechanical Engineering
Wesley Chan, Ph.D. Student, University of Tokyo
Elizabeth Croft, Professor, UBC Mechanical Engineering
B.2 Consent

HRI-Cues: Human-Robot Handover Study Consent Form

Project Title: HRI-Cues: Human-Human Handover Study II
Principal Investigator: Dr. Elizabeth Croft, Wesley Chan and Matthew Pan
Co-Investigators and Contact Persons: Wesley Chan and Matthew Pan
Funding: This research is funded by the National Sciences and Engineering Research Council of Canada (NSERC).
Purpose: The purpose of this study is to observe and kinematic behaviours in human-human handovers, and to explore how a human giver and receiver negotiate handovers of several objects. Results from this study will be used in subsequent research to improve the ability of robotic assistants to interact with non-expert human users.
Procedures: Before the actual handover experiment, you will be asked to fill out a preliminary questionnaire which asks you for some demographic information (e.g., age, gender, etc). For the experiment, you will be paired with another participant to perform a series of handovers using several everyday objects while wearing a
jacket and a cap with motion capture markers. One participant will play the role of the giver, while the other will be the receiver. During the experiment, both the giver and the receiver will be standing. Both the giver and receiver will start with their hands placed at their side. On a verbal Go signal, the giver will retrieve an object, and hand it over to the giver. After object transfer, both the giver and the receiver will return to their start positions.

The experiment will last approximately an hour. You may refuse to participate in this experiment and you may withdraw at any time. We will be recording motion capture and video data, although the latter is not required for your participation.

**Potential Risks:** Slight temporary fatigue from passing various objects back and forth.

**Confidentiality:** No identifying information will be collected or stored with your data. Data collected from the survey will be stored on a password protected computer or a locked cabinet in the CARIS Lab, which has restricted secure access and is locked at all times. If you have any questions or concerns about what we are asking of you, please contact the study leader or one of the study staff. The names and telephone numbers are listed at the top of the first page of this form. If you have any concerns about your rights as a research subject and/or your experiences while participating in this study, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

**Consent:** By signing this form, you consent to participate in this study, and acknowledge you have received a copy of this consent form.

I agree to allow myself videotaped during this experiment (please circle one):

**YES**  **NO**

Name (print): _______________________________ Date: __________

Signature: _________________________________

Last Revised: November 3, 2014  Rev 1
Appendix C

Supporting Materials for Investigating Human-to-Robot Handovers

C.1 Consent

Consent for Participation in Research

This is an informed consent ("Consent") provided by the individual identified below who has agreed to participate in a research study ("Study") conducted by Disney Research.

Study Title: A study of proprioceptive and kinodynamic behaviors during human-robot interaction with visual and haptic feedback

Principal Investigator: Matthew Pan, Lab Associate
Address/Contact Info: Disney Research Los Angeles, 521 Circle Seven Drive
Los Angeles CA 91201
E-mail: [redacted]
Phone: [redacted]

Senior Advisors: Günter Niemeyer, Senior Research Scientist; Lanny Smoot, Disney Research Fellow

Other Investigator(s): Vinay Chawda, Postdoctoral Fellow

Purpose of this Study
The purpose of this Study is to observe user kinematic, proprioceptive and haptic behaviors in human-robot and human-human handovers of objects with and without the user simultaneously immersed in a virtual reality environment via a head-mounted display. Results from this study will be used in subsequent research to improve the ability of robotic assistants to interact with non-expert human users during co-operative object manipulation and enhancing virtual reality experiences with robotics-based haptic interactions.

Procedures for Study:
Before the Study, you will be asked to fill out a preliminary questionnaire which asks you for some demographic information (e.g., age, gender, etc.). You will perform a series of object manipulation tasks while seated including, but not limited to:

- Picking up and putting down objects
- Handovers and/or toss-catch exercises with a robot or an experimenter
- Juggling, or attempting to juggle balls

Depending on the experimental condition, you may or may not be immersed in a virtual reality environment which will be rendering the physical objects that are to be handed over or tossed by an experimenter or robot. In each trial, you will be asked to evaluate the interaction. For the purposes of tracking your motions within
the virtual environment, you may be asked to wear a jacket, cap and/or pair of
gloves with motion markers. Following the Study, there will be a debriefing survey
to collect feedback on your experience during the Study.

**Duration and Location of Study:** 1 Hour, Disney Research Los Angeles

**Participant Requirements:**
To participate in the Study, you must:

- Be over 18

- Never have had a seizure or blackout before and have no history of seizures or epilepsy

**Risks**
During the Study, you will be donning a virtual reality head-mounted display that will fully occlude your vision of the physical environment. You may also be performing interaction tasks (e.g., object handovers) with a robot arm. To minimize the risks of tripping/falling/collisions, you will be seated at all times when interacting with the robot and wearing the head mounted display. Additionally, objects used will be lightweight. Other risks of participation include: motion sickness/-nausea, eye-strain, fatigue and repetitive stress injuries of muscles, joints and skin, tripping and dizziness. To reduce these risks, you will be able to take a break or discontinue participation in the Study at any time.

**Benefits**
There may be no personal benefit for you from your participation in the Study but the knowledge received may be helpful to this research and may be of value to humanity.

**Compensation & Costs**
There is no compensation provided for this study.

**Privacy Policy/Ownership**
The security, integrity and confidentiality of your information are extremely important to us. We have implemented technical, administrative and physical security measures that are designed to protect guest information from unauthorized access, disclosure, use and modification. From time to time, we review our security procedures to consider appropriate new technology and methods. Please be aware that, despite our best efforts, no security measures are perfect or impenetrable.

The Study and its data, information and results as well as all ideas and suggestions communicated by you during your participation in the Study, will be owned exclusively by Disney Research, and may be used, published and/or disclosed by Disney Research to others outside of Disney Research. However, your name, address, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by Disney Research. You will not receive credit in any materials in connection with the Study.

Permission to Use Recorded Audio Visual Data
Disney Research may want to record and use a portion of any video, audio or other recording for illustrative reasons in presentations of this Study for scientific or educational purposes. Please initial below if you wish to consent to the following uses of your recorded data:

I give my permission to the following uses of my recorded data:

1. Use of my recorded data in research venues, including publications & public presentations, for illustrative and educational purposes:
   Please initial here: _____YES _____NO

2. Use of my recorded data without additional identifying information (your name, address, etc) by other research groups or for other research purposes than those described herein in the future:
   Please initial here: _____YES _____NO
If you give your permission to any or all of the uses above, this means that you have granted Disney Research the right to use and modify your recorded data as required by Disney Research for any reason related to the permitted uses of the data.

**Rights**

Your participation in the Study is voluntary. You are free to stop your participation in the Study at any point. Refusal to participate or withdrawal of your consent or discontinued participation in the Study will not result in any penalty or loss of benefits or rights to which you might otherwise be entitled. The Principal Investigator may at his/her discretion remove you from the Study for any of a number of reasons. In such an event, you will not suffer any penalty or lose any benefits to which you might otherwise be entitled.

This Consent form applies only to the Study described above and does not modify or replace any other agreements you may have with Disney Research or The Walt Disney Company.

**Right to Ask Questions & Contact Information**

If you have any questions about this Study, you should feel free to ask them now. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by mail, phone or e-mail in accordance with the contact information listed on the first page of this Consent.

If you have questions pertaining to your rights as a research participant; or to report concerns regarding this Study, you should contact the Chair of the IRB Committee, Steve Stroessner.

Email: [email]
Phone: [phone]

**Voluntary Consent**

**ADULT PARTICIPANT**

By signing below, I agree that the above information has been explained to me and all my current questions have been answered. I have been encouraged to ask ques-
tions about any aspect of this research Study during the course of the Study and in the future. By signing this form, I acknowledge that I have read and I understand this Consent, this Study has been explained to me and I agree to participate in this research Study.

__________________________    _________________________
PARTICIPANT SIGNATURE        DATE

__________________________    _________________________
Print Name                  Telephone

__________________________    _________________________
Street Address                City, Sate, Zip

Email Address

I certify that I have explained the nature and purpose of this research Study to the above individual and I have discussed the potential benefits and possible risks of participation in the Study. Any questions the individual has about this Study have been answered and any future questions will be answered as they arise.

__________________________    _________________________
SIGNATURE OF PERSON OBTAINING CONSENT    DATE

Last Revised: November 14, 2016
### C.2 Survey

Participant: ________________

**Condition (circle one):** 1 2 3 4 5 6 7 8

Using the scale provided, how closely are the following attributes associated with the robotic handover you just performed (Place a checkmark in the circle)?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A moderate amount</th>
<th>Very much so</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dexterous</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awkward</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competent</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociable</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scary</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledgeable</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clumsy</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strange</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsive</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awful</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassionate</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dangerous</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capable</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeling</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trustworthy</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with fragile objects</td>
<td>o o o o o o o</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Last Revised: April 2, 2017

Rev 2