Title
Enabling Assistive Service Robots to Contextually Organize Household Objects

Author
Adeleye, Akanimoh Oluwasanmi

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Enabling Assistive Service Robots to Contextually Organize Household Objects

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science

by

Akanimoh Oluwasanmi Adeleye

Committee in charge:

Professor Henrik Christensen, Chair
Professor Sicun Gao
Professor Nadapa Nakashole
Professor Michael Yip

2023

The Dissertation of Akanimoh Oluwasanmi Adeleye is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my friends, to my family, to my loving girlfriend. All of whom I would not want to do this without.

And to all my ancestors and the ancestors of those who have struggled for freedom and a better opportunity at life for themselves and their family.

EPIGRAPH

If there is no struggle, there is no progress

—Frederick Douglass

Our greatest glory is not in never failing,
but in rising up every time we fail.

—Ralph Waldo Emerson

But man is not made for defeat.
A man can be destroyed but not defeated.

—Ernest Hemingway

I continuously remind myself that no one is guaranteed an easy life,
but we can make it easier for others

—Frances Arnold

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

May this work enlighten, inspire, and help further understanding of the needs and capabilities of robotics in household environments.

Nothing that is created was not first desired. Although this is not true for all human creations, inventions are the representation of a desire and idea of someone or some people.

ACKNOWLEDGEMENTS

I would like to first acknowledge and thank my advisor, Professor Henrik Christensen as the chair of my committee. Your support to my education and personal development is has been conspicuous in our one on one meetings and your handling of various academic circumstances through out my journey. You gave me the freedom to explore my research interest, with safety guards, and have always been considered about Sanmi the researcher, and Sanmi the person. Thank you for that.

I would also like to thank other professors who have help mold me and and shaped my experience towards getting a PhD and research at large. First to professor Sesh Commuri, you allowed me to join your lab at the University of Oklahoma during research with PID controllers. This was my first experience of grad school and the mentorship, and thrill of discovery in your lab encouraged me to continue. To professor Cindy Grimm, my summer research experience with you over the summer was truly live changing. Thank you to you and Bill for hosting research experiences at Oregon State University. Your care and mentorship has extended well beyond that summer and you have always supported me however you can. You have shown me how to do research in a thoughtful way and how to balance this with a fun life filled with adventures and hobbies. To Rachel Burcin, I am so glad you were the director of the summer research program at Carnegie Mellon University. I would not have wanted to met any other director. You move fast, as you understand that life moves fast, and you pushed me to the same in a positive way. To Chad Jenkins, thank you for being an advocate and providing a space for me and the students at Michigan University as well as across different Universities. To Laurel Riek, I enjoyed the time I spent in your lab my first few years of grad school. You provided me a space with deep research support and also changed my presentation and writing style for the better. And finally, to a myriad of other professor, Nikolay A. Atanasov, Sicun Gao, Julian Mcauley, Micheal Yip, and Nadapa Nakhaol, thank you for your interest in my research and supporting me and my growth educationally as a researcher.

To my wonder lab mates, thank you for our time working late doing research together

and learning form each other: Dr. Shengye Wang, Dr. Priyam Parashar, Dr. Ruffin White, Andi Frank, Jiaming Hu, Shixin Li, Michelle Sit, Quan Vuong, Anwesan Pal, David Paz, Yiding Qiu, Christopher D'Ambrosia, Abdulaziz Almuzairee, Alexander J Langley, Hengyuan Zhang,Srinidhi Bharadwaj, Seth Farell, Shrutheesh R Iyer, Jing-yan Liao, Luobin Wang, and Narayanan E. Ranganatha. Thank you for helping grow as a scientist and for many memorable moments these past six years.

To my friends. Jeffery Wang, I am beyond happy to have met you during my PhD. Our friendship was solidified through the rigours classwork of UCSD and briefly working in the same lab. Our conversations life and research lead to some of the exploration in my work and continue to inspire me in how I think. Florian Richter, you are an amazing researcher and friend. I am also extraordinary gratefully for your input on any topic and energized by your positive approach at life. Pamela Ortega, moving midway across the country for graduate studies was challenging. I consider myself very fortune that a friend from Oklahoma was also accepted and came to UCSD for their graduate studies. Thank you for helping to edit my papers, and for your friendship. To Shady Ghoneim, Caitlin Clement, and Teresa Ayala, our trip to Egypt is one of my favorite memories with you all. These moments and memories, along with phone calls and check-ins, have helped supported me throughout grad school. To my research experience for undergraduate friends: Eley NG, Nick Bira, Autumn Gael, I consider it so lucky that we could all met. I am proud to be your friend and proud that we all continued on towards our doctorate degrees. To Jay Patravali and Liz Olson. Our summer at CMU feels as fresh in my mind and yesterday. Both of you encouraged and inspired me towards grad school. Finally, to a myriad of other friends, Ariana Mirian,Sophia Krause-Levy,Tricia Light,Ben Shih,Jessica Sandoval,Adriane Minori,Kelechi Agwuncha,Taylor McKie, Tyler Sanders,Grace Dunbar, and Sagiv Levi, thank you for your friendship, adventures, and talks about life and research. Lastly, to Blanca Fernandez. Meeting you in grad school as been one of the greatest joys of my life. You have been with me through the latter half of my PhD. You have seen my low periods during this journey and have helped make space for them. You have seen my high periods and have

celebrated them even more than I have at times. I am eternally grateful to see you model what a truly positive mindset can look like and how to prioritize happiness and joy. These things are important in life and have been a pillar of support through grad school.

For research in Chapter 3 of this dissertation, I thank Jiaming Hu for his assistance with pose estimation of objects. This chapter contains material from ”Putting away the groceries with precise semantic placements.” by Akanimoh Adeleye, Jiaming Hu, and Henrik I Christensen, which appears in Proceedings of IEEE 18th International Conference on Automation Science and Engineering (CASE), 2022. The dissertation author was the primary investigator and first author of this work.

For research in Chapter 4 of this dissertation, I thank Shrutheesh R. Iyer for his work on system integration and control architecture; Anwesan Pal for his work on navigation; Jiaming Hu for his work on robotic manipulation; and Aditya Aggarwal for his work on object and receptacle detection. This chapter contains material from ”Household navigation and manipulation for everyday object rearrangement tasks.” by Shrutheesh R. Iyer, Anwesan Pal, Jiaming Hu, Akanimoh Adeleye, Aditya Aggarwal and Henrik I Christensen, In Submission for the 7th IEEE International Conference on Robotic Computing (IRC), 2023. This dissertation author was one of the primary investigators and authors of this work.

For research in Appendix A of this dissertation, I thank Auriel Washburn for her mentorship and stewardship on the project and Thomas An for his assistance with robot system development. This chapter contains material form ”Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust” by A. Washburn,A. Adeleye,T. An and L.D Riek, In Proceedings of ACM Transactions on Human Robot Interaction (THRI), 2020. This dissertation author was one of the primary investigators and authors of this work.

VITA

2015	Undergraduate Student Researcher, Oregon State University
2016	Undergraduate Student Researcher, Carnegie Mellon University
2017	Bachelor of Science, University of Oklahoma
2017–2023	Graduate Student Researcher, University of California San Diego
2021	Teaching Assistant, Department of Computer Science and Engineering University of California San Diego
2017,2022	Internships: Intel, Apple.
2020	Masters of Science, University of California San Diego
2023	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Akanimoh Adeleye, Jiaming Hu, and Henrik I Christensen, ”Putting away the groceries with precise semantic placements.” In Proceedings of 18th IEEE International Conference on Automation Science and Engineering (CASE), Aug 2022.

Shrutheesh R. Iyer, Anwesan Pal, Jiaming Hu, Akanimoh Adeleye, Aditya Aggarwal and Henrik I Christensen, ”Household navigation and manipulation for everyday object rearrangement tasks.” In submission of 7th IEEE International Conference on Robotic Computing (IRC), 2023.

Washburn, A., Adeleye, A., An, T., and Riek, L.D. (2020). ”Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust”. In Proceedings of ACM Transactions on Human Robot Interaction (THRI)

Akanimoh Adeleye, Jiaming Hu, and Henrik I Christensen, “Place-and-pick-based re-grasping using unstable placement,” in ISRR - Robotics Research, (Geneva, CH), IFRR, Springer Verlag, Sep 2022.

ABSTRACT OF THE DISSERTATION

Enabling Assistive Service Robots to Contextually Organize Household Objects

by

Akanimoh Oluwasanmi Adeleye

Doctor of Philosophy in Computer Science

University of California San Diego, 2023

Professor Henrik Christensen, Chair

Service robots are designed to assist and aid humans in the day to day tasks as well as specific and minute tasks. In recent years there has been an increase in both demand and production of service robots. These robots are slowly becoming more prolific both in personal settings such as the home, and professional settings such as restaurants and hospitals. Despite this growth, there are still major challenges preventing service robots from being ubiquitous and operating intelligently in these unstructured domestic environments.

Like many autonomous robots, the technical challenges of service robots revolve around three main facets: sensing and modeling their environment, understanding and reasoning about said environment, and acting in a way that is safe and conducive towards a task goal. Unlike

industry robots, service robots operate in dynamic environments and have tasks that cannot always be completed repetitively. This presents a rich set of challenges to solve in all three areas.

One exemplary task of this for service robots is tidying objects within the home. This task is important as it helps assist with activity of daily living (ADL) while meeting consumer demand. A few concrete illustrations of the challenges of this task are shown when we consider the robot must sense objects and ground language of these objects before we can determine where the objects belong. Notions of a correct placement vary as each user has their own personal preference to organization. Grasping and manipulating the objects as well as navigating within the environment must also be accomplished in order to successfully complete the task.

In this dissertation we present several contributions that help address these challenges and the surrounding gaps towards solving these issues. First, we explore the use of a service robot to put away the groceries in a precise and stable configuration that satisfies the desired semantic relationship of grocery and pantry objects. Second we explore the ability of our robot to put away objects throughout a home environment based on a user preference using a recommender system approach. We share details of our implementation of these systems through real-world experiments.

Chapter 1

Introduction

Robotics is a ever growing field in terms of research, units of manufacturing and sales, as well as the public imagination of what is capable for robots. Although the first and second are only recently true relative to the history of people, robots have always fascinated people. Specifically, the type of robots that resemble humans and other living things. Since 350BC, automata like that created by Archytas - an mechanical bird called ”The Pigeon”; and a bell striking water clocks, using human figurines - made by Harun al-Rashid, have shown a desire to create machines that imitate life and can replace humans in working tasks[119, 113]. Part of this fascination comes from a desire for companionship and an innate curiosity or desire to understand life and create or mimic life ourselves. This is partially why robots are easily anthropomorphized, so long as the ”Uncanny Valley” is avoided, and are extraordinarily suited to become mechanisms for human companionship and care-giving. This is the main essence of service robotics: to preform useful task that are beneficial to people.

Today’s desire for service robotics are not much different than the pining of previous groups. Despite common fears of new technological and robotic take overs [102], increased consumer demand for service robots world wide [4, 115] shows that this desired is still held by society at large. Unlike humanoid robots, most service robots do not explicitly try to mimic or resemble human physiology and interaction, but do so to some degree as a bi-product of their design and interaction with people. So while service robots are still easily anthropomorphized, they often do not resemble humans enough to fall into the ”Uncanny Valley”. This helps negate some of the negative emotions that may come with their integration to society. We are now much closer than we have ever been before to building life-like mechanises capable of intelligently preforming multiply general tasks useful to people. Although this is still quite far away, gradually we find the current capabilities of service robots are allowing them to be directly integrated to society.

Robots in society are changing our lives and we are witnessing service robots showing great promise for supporting people. In both domestic and professional settings service robots are performing tasks like supporting workers and users by cleaning and tidying up [150, 6, 80], assisting the elderly and handicapped [52, 63, 87], preforming inspections and maintenance [170, 39], enhancing surgical operations and agricultural needs [134], and exploring dangerous or unreachable environments [46, 147]. Despite this progress, there are still many challenges to increase the current capabilities of service robotics. This creates a gap that prevents them form being more ubiquitously integrated in society. For personal assistive service robots, we still do not see them integrated into society unless they preform specialised tasks such as vacuuming or are social companions. This is very different from the envisioned idea of service robots shown in modern mass media. Albeit robots do not have to adhere to every vision of the masses, such as in the case of flying cars versus airplanes, it is clear there is a disconnect from the type of support people envision from robots and what they currently can provide.

My work is situated at the intersection of this problem and focuses on increasing robots capability to preform assistive tidy up tasks in dynamic real-world environments such as the home. This work enables robots to preform these tasks by combining expert domain knowledge with personal user preference data, along with integrating this knowledge with advances techniques in object detection, navigation, and manipulation.

1.1 Motivation

The primary motivation for this work is based on direct application use cases and a general desire to test, evaluate, and improve the capabilities of service robots in the real world. Specifically, we look at service robots that must assist users in rearranging objects and tidying up the house.

Due to the many challenges in this field and the expansive nature of this task, previous works have explored various aspects of this problem. Some work focuses on the system integration of current robotic skills and increasing reliability to allow service robots to operate and complete complex cleaning tasks for long periods of time unhindered[174]. Other work focuses on the high level reasoning, semantic understanding, and decision making needed when executing such complex tasks[50]. Lastly, some works focus on improving capability of specific tidy up tasks such as folding laundry, loading and unloading the dishwasher, as well as cleaning and picking up items [82, 7, 37, 70]. There are numerous other works that focus on improving general robot capability such as grasping and perception, that also help enable robots to complete tidy up tasks more efficiently as well.

In this work, we also aim to further close the gap in robotic capabilities and their use in the real world. However, we take a different approach than some of the previous work mentioned above. While training in simulation allows researchers to abstract from the technical skills needed for a real word task and focus on the high level reasoning needed for such skills, this often presents a problem when transferring to the real world. On the contrary, work that focuses purely on increasing the skills for such task but not the high level reasoning nor how to personalize that reasoning to each user, only partially solves the problem.

The aim of this work is to explore both the technical skills needed for a real world task and how to personalize high level reasoning needed for tidying up to each user. Additionally, since these robots will be deployed in real homes with real people, we also aim to use deep learning methods in an explainable way.

1.2 Thesis Problem

The use of service robots to assist users and complete rearrangement and tidy up tasks poses significant challenges. Therefore, in this dissertation the following hypothesis is tested:

> The use of contextual knowledge allows service robots to both assist users towards, and efficiently complete, the task of organizing household objects while solving part of the general problem of robotic household organization.
>
> This hypothesis was tested via the following methods:

1. Using a service robot to put away grocery items in a kitchen pantry:
   - Design and construction of contextual knowledge formalized and integrated as part of high-level reasoning.
2. Using a service robot to put away household objects throughout a home environment.
   - Exploration of combining user preference data with generative Large Language models for tidying up task.
3. Integration of contextual knowledge and the technical skills needed for a service robot towards tidy up task.

1.3 Scope

Robotics is a large domain involving multiple disciplines:Computer Science, Electrical Engineering, Mechanical Engineering, Biology, Statistics, etc. This is due to the complexity of robotics as well as the vast range of what is considered a robot. Because of this there are many different areas of active research in robotics, ranging from different types of industrial and service robotics, to medical and soft robotics. This dissertation focuses specifically on personal or domestic autonomous service robots and how to use context to enable them to complete rearrangement and tidy up tasks.

1.3.1 Artificial Intelligent Personal Service Robots

A personal service robot is defined as a robot that performs useful tasks for humans or equipment in non-commercial applications[115]. These robots therefore primarily operate in unstructured domestic environments such as the home. The use of context can help robots be flexible and still robust within the uncertainties of such environments.

The notion of context has been explored across various disciplines such as Cognitive Science, Psychology, and Artificial Intelligence(AI). Within the realm of robotics however, a definitive and unified definition of context remains elusive. Typically definitions of context given in AI have been adapted for use in robotics. This initial definition from AI focused on " how knowledge can be represented symbolically and manipulated in an automated way by reasoning programs"[94]. In application, this definition is heavily reliant on semantic and symbolic representations of knowledge, grounded as formal logical expressions. Through predicates that describe concepts about the world and actions the robot can take, which are connected by expressions, robots engage in reasoning to predict the outcomes of their actions.

Recently, the definition of context and knowledge representation in AI has been expanded explicitly for robotic systems in [124]. Here, knowledge representation is defined as "a means of representing knowledge about a robot's action and environment, as well as relating the semantics of these concepts to its own internal components for problem solving through reasoning and inference." Knowledge representation can be viewed as high-level or low-level representation. High-level representation involves structured, semantic descriptions of relationships among different components, abstracting information into meaningful concepts that facilitate problem-solving and reasoning; low-level representation would be focused on specific technical data, such as sensor readings or control parameters. Both provide context to the robot, influencing how it perceives, understands, and interacts with its environment. There are many examples and methods of how to include context within a robotic system that are elaborated more in [124].

In general we can think of applications of context in robotics as what information can we give to the robot, what information can the robot gather from the environment autonomously, and what reasoning algorithms can we construct to allow the robot to effectively utilize the information it is given and gathers from the environment for task related goals. Simply stated, context as a high-level knowledge representation, gives meaning to the input a robot acquires and uses in task.

1.3.2 Learning and Large Language Models

Learning is used extensively through service robotics. While learning is closely related to context, it differs in that learning is not succinct enough alone to be a representation [121]. While this was true in the past, with the advent of Large Language Models(LLMs), capably of capturing long-rage dependencies, this line is becoming increasingly blurred.

Regardless, machine learning and deep learning methods, such large language models, help enable robots to preform complex task on par to and sometimes surpassing human ability, by using data patterns to understand, reason, classify and inform robotic actions and decisions. These techniques make robots more intelligent, efficient and able to adapt to challenging task and environments [160]. There are many important examples and seminal work for robotics with learning used as the for front: object detection and recognition, speech generation and understanding, motion planning and control, localization, etc. In this work, we use learning techniques and try to do so in an explainable way.

1.3.3 Robotic Applications

When service robotics was still in its nascent stage, a noticeable distinction from industrial robots of that era was evident. Service robots were envisioned not only as tools that operated for humans, but also as possible collaborative machines working alongside them to accomplish tasks. Early glimpses of service robotics applications can be found in [32], where the potential of robots to aid patients in medical facilities and other institutions, provide assistance to individuals with disabilities and the elderly in domestic settings, and even integrate into wheelchairs as robotic arms were explored. Since then, significant strides have been undertaken to fully realize these applications; however, they continue to linger on the horizon, and have not yet fully been achieved. In parallel, other applications for service robotics have surged forward, experiencing remarkable growth. While some of these new applications have moved closer to realization, others still remain tantalizingly yet to be fully realized. In this section, we cover some current applications for personal service robotics.

When considering what applications for personal service robots users care most about, [78] explores surveys that have asked people about the tasks that personal service robots should perform. They list three primary consensuses that users want:

- Personal service robots that are capable of performing complex tasks such as cleaning the house, washing dishes, doing laundry.
- Personal service robots that can perform a wide variety of tasks: tidying the house, guarding, and helping children with their school work.
- Personal service robots that can complete a task on their own but also collaborate with users to accomplish a task. This was especially true for those with disabilities.

Looking at the type of personal service robots currently available [63] categorizes the abilities and assistance of personal service robots into four categories:

- Physically assistive service robots can aid in physical task such as vacuuming and mowing. Robotic wheelchairs and manipulators for lifting, personal care and other physical acts of daily living are other examples.
- Sensory assistive robots such as robotic guides for the blind and telepresence robots can provide aid to users by supporting increased mobility and expanding sensory ability.
- Socially assistive service can provide aid by acting as companions and toys, often helping with cognitive or social task.

- Mixed assistive service robots can provide aid often using any two or all three of the aforementioned categories. Physically and socially are the most common as in the case with robots designed for the elderly.

Significant overlap between the current applications of service robots and the wants of users indicate that while personal service robotics is working on the correct task, the field as a whole still lacks the skill and capabilities to fully autonomously complete most assistive task. In this dissertation, I focus on physically assistive personal service robots that can complete users wants for performing organizational task.

1.4 Contributions

The contributions of this work are as follows:

- **Present a methodology for a service robot to put away the groceries in precise configurations that satisfies desired semantic relationships [7].** In this study I gather and use food ontology data, along with common geometric container packaging constraints and semantic word association, as contextually knowledge to organize new grocery items in a users pantry. I show how these forms of knowledge can be given a mathematical representation using hierarchical encoded vectors. Incorporating this representation, I present a function to compare the similarity between two food items. Finally I show through real world experimentation a service robot organizing grocery items in different existing pantry configurations.

- **Design and present a methodology for a service robot to put away the house-hold object throughout a home environment according to users preference [76].** In this study I explore the use of explicit user object location preferences data using collaborative filtering as a recommender system based approach. I augment this approach to incorporate a Large Language Model using the embedding space of the model. Finally, using the inferred preferences of our users, we demonstrate a robot system capable of navigating from room to room while identifying misplaced objects and returning them to their correct location.

- **Real world application of contextual knowledge and the technical skills need for service robots in tidy up task.** Through both of the contributions above I show that it is possible for a real world robot to accomplish organizational tasks efficiently. I highlight the specific computer vision, motion planning and task level planning skills needed for such organization task as well as future improvements that are needed these skills and for the overall integration of these skills. These improvements look towards the future of enabling robots to preform organization task as products used by consumers in real homes.

1.5 Dissertation Outline

- **Chapter 2** provides a brief overview of the related work in the areas of service robotics, both their over all skill development and their application towards home organization task.

- **Chapter 3** presents new methodology for using contextual knowledge's to enable a service robot to organize grocery items in a home pantry.

- **Chapter 4** describes the use of personal user data augmented with general large language models to organize house hold objects in a home setting.

- **Chapter 5** summarizes the main contribution of this dissertation, providing a discussion on open questions and future work as well as closing remarks.

Chapter 2

Background

This dissertation, focuses on improving the capabilities of personal service robotics. As mentioned in Chapter 1, at its core, a service robot is a type of robot that performs useful tasks for the benefit of humans or equipment. The development of service robots has grown prodigiously over the past fifty plus years. This section provides information about the field of service robotics and the related works of this field.

2.1 Early beginnings: Classical Service-Robots

A few examples of early service robots can be shown in [22, 148, 114, 107, 161, 183, 77]. These robots were the beginning of both autonomous mobile robots as well as robots with some of the skills needed for service robotics: manipulation and grasping of novel objects, navigation and obstacle avoidance, as well as task and motion planning.

Shakey [22], built in 1966, is accredited as being the first general purpose robot. It's architecture and algorithms became blueprints for some service robotics. The integration of camera, optical range finder, bump detectors and other sensory components, along with the introduction of the STRIPS planner and A* search algorithm, made Shakey extraordinary novel. After Shakey, robots such as SAGE [114] and RHINO [19] soon followed. Similar to Shakey, albeit more advanced, these robots were equipped with early sensor capabilities and were able to perform navigation and obstacle avoidance as a result. Both robots were used as tour guide robots and impressively, SAGE showed signs of reliability in long-term autonomy for service robots as it was able to go a total of 135 error-free days.

In the late 1990s and early 2000s, we started to see the rise of robots entering the personal sector. Robots such as the Care-O-bot II [52] and Nursebots [107] were not only designed to be placed in homes and assisted living facilities as research objectives, but were also deployed in these environments to explore commercial aims. Nursebot provide aid to elderly users by autonomously guiding users in the facility. Care-O-bot II, unlike the robots mentioned so far - and its predecessor Care-O-bot I, was a mobile platform with a manipulator arm. Although the addition of manipulators on mobile bases had already been researched [15, 31], Care-o-bot II was one of the first to do so in the hopes of a commercial product rather than a prototype. Excitingly, this meant Care-o-bot II - in addition to being a walking support robot - was also able to execute simple fetch and carry tasks and had early speech recognition software. This time period was also the birth of the titan of personal service robots: the Roomba. The Roomba first appeared in 2002 and would soon lead the charge for robot vacuums to be the leading type of personal service robot to date[4].

With robots in a position where commercial personal usage was no longer something envisioned only within mass media but rather an emerging reality, research on service robots was gearing up. More robots were developed to be deployed and tested in real world environments. These robots varied in ability and design but a few notable algorithms and techniques emerged during this time. First, early mapping, localization, and planning techniques were some of the hallmarks of these era. These classical works primarily solve the task and motion planning problem through symbolic planning such as STRIPS, SHOP, and HTN [45, 111, 43]. Famous algorithms and techniques such as the A* search algorithm, early Rapidly Exploring Random Tree (RRT) methods[18], Probabilistic Road Map (PRM) [81] and the use of potential fields [139], were developed or modified to fit robotic planning use. Seminal early Simultaneous Localization and Mapping (SLAM) algorithms were also just beginning to show teeth [93, 158].

During this time period we also saw early approaches to grounding language and autonomous grasping. Early work on autonomously grasping objects used pre-stored grasp primitives as well as techniques based on friction cones and form- and force-closer, or other similar methods, to grasp objects with known 3D models [100, 14, 104]. For objects not known, methods to determine the 3D model of the objects were applied but had difficultly obtaining full and accurate 3-d reconstruction due to the state of optical sensors and techniques at the time. Some early methods also tried grasping unknown objects without relying on 3-d reconstructions, but instead learning where to grasp based on sensory input of the objects [148]. A slew of early approaches to grounding language were developed during this time. Work by Roy and colleagues [69, 144] give a good overview of the many efforts towards this task. Towards the end of this period, we began to see a rise in using machine leaning as a process to relate language to visual perceptions and physical actions.

The skills detailed within this section encompass foundational elements of robotics and can be applied beyond personal service robots to many other robotic applications. This in part highlights the complexity and challenge of building personal and professional service robotics.

2.2 A Modern take: Skilled Service-Robots

Due to the interdisciplinary nature of robotics, as the fields surrounding robotics improved, so to did robotics. Modern advancements in computing power, computer vision and perception, and early deep learning techniques immensely improved the capabilities and skills of classical robotics.

This significant transformation became particularly evident in the middle to late 2010's when we saw a convergence of improved and functionally reliable hardware, with smarter software, that paved the way for more commercially ready robots. Progress in sensor technology, including Lidar, 2-D and 3-D cameras, alongside advancements in computer vision, enabled robots to enhance their visual perception and modeling of the world. Meanwhile, innovations in mechanical engineering and conscientious robotic design led to enhanced actuators, resulting in sleeker, more compact robots that are safer to coexist with humans and easier to integrate into homes. Simultaneously, advancements in computing power and chip memory facilitated the deployment of smaller, cost-effective onboard computers, enabling the execution of more sophisticated, memory-intensive algorithms in real-time. These collective improvements opened the door to a plethora of new research aimed at continued enhancement of robot capabilities.

Challenges such as the 2013 DARPA Robotics Challenge helped showcase and also advance the current status of robotic technology and research, particularly in the areas of mobility, dexterity, manipulation, perception, and operator control. And in 2017 we saw the explosion of robots entering professional and personal spaces. Robots were now starting to become prolific in consumer and commercial spaces as the cost for robot components were decreasing while their capability increased significantly. The Roomba now had over 10 million robots in homes [1] and more and more commercial robots such as, food delivery robots and professional cleaning robots, were on the rise. The year 2017 was deemed the year robots went everywhere [157].

Continued technological progress and the development of specific robotic skills fostered comprehensive advancements in service robots during this period. Work by Lee [90], underscores three key technological foundations for comprehensive advancement in both professional and personal service robots:

- Fetch, Detection, and Navigation: This foundation encompasses improvements in object retrieval and manipulation, environmental sensing, and navigation capabilities.
- Human-Robot Interaction: This domain emphasizes advancements in how humans and robots interact, encompassing communication and cooperation.
- Architecture and Platform: Here, the emphasis lies on the development of robust underlying structures and platforms that enable the effective integration of these technologies into service robots.

A few examples of modern service robots form this area, exemplifying these advancements, are shown in [161, 183, 77, 99, 82, 8, 7, 70, 125]. Readers are encouraged to read [185] and [90] for further reading on past and current service robotics research.

2.3 Larger capabilities: Robots that Reason and Communicate

Looking to the future, it is no exaggeration to believe that Robotics is once again at an inflection point. Almost fittingly fortuitous, the catalyst for this inflection point arrived in the same year as the previous turning point: 2017. In that year, the now-famous paper ”Attention is all you need” was published [171] and introduced Large Language Models(LLM) as well as the Transformers Architecture. Similar to the previous inflection point, where it took time for robotics algorithms to align with technological advancements, the effects of this paper took time to manifest in robotics.

The capability to ground language is not a new skill or field of study, but the contextual understanding Large Language Models(LLM) have shown for language in long-range dependencies and relationships, make them able to ground language better in complex context. The transformers architecture has provided is a mechanism to process sequences of data in parallel rather than sequentially. Unlike previous attempts to capture long-rage dependencies in data efficiently, such as RNNs [187], the the transformer architecture is able to do so while being flexible and still scalable.

New research efforts in robotics exploring the capabilities of Large Language Models are emerging [182, 75, 152, 163]. Their impact on the field of robotics is already evident and will be more profoundly felt in the upcoming years.

2.4 Service Robots in the Home: Tidy-up

The sections above gave a overview of some of the major developments of service robot. In this section, we magnify our understanding of previous research efforts in service robotics to a single application environment: robots in the home, commonly refereed to as personal or domestic service robotics. We are are particularly interested in robots that attempt the tidy-up task [9].

The motivation behind personal service robots that can help clean and tidy a home is explained in Chapter 1. As cost for service robots continue to decreases, and the general development of robotics skills increase, we will eventually see personal service robots in homes with the ability to clean and tidy houses. Research towards this goal has been ongoing for as long as service robots have existed.

Early work in [122, 28] conduct user surveys where they show the importance of tidy up task and find that users elicit strong feelings with respect to organization of personal items, indicating users preference must be considered in such task. Later work focuses on enabling the capability for service robots to complete organization task, such as understanding organizational structures through ontologies and user preferences data [150, 6, 80], grasping of objects [99, 70], grounding of semantic labels to room locations [121], learning from demonstration and end to end task completion [77, 184], and most recently, the inclusion of LLMs [79, 182, 75]

2.5 Chapter Summary

Robots capable of supporting and replacing humans in various tasks have long been a desired goal throughout history by many communities. Over the past fifty years, service robots have experienced remarkable growth, moving closer to fulfilling this desire for more complex tasks in both professional and personal domains. This chapter delves into the history and evolution of service robotics, with a particular emphasis on its growth over the last half-century. A central focus lies in the effort to enhance the capabilities of these robots in domestic chores, specifically cleaning and tidying.

At their core, these improvements revolve around enhanced sensing, the development of better environmental models, and the ability to reason and understand the world in which the robots operate. Moreover, the capability to interact with humans and objects to execute actions reliably is fundamental to the essence of service robotics. While a service robot lacking some degree of mastery in these skills might accomplish certain tasks, it is doubtful that such a robot can operate autonomously effectively or efficiently in real, unstructured environments alongside humans performing multiple tasks. Building a service robot that can achieve this demands a mastery of all these skills, making service robotics a challenging endeavor.

Chapter 3

Putting away the Groceries

We present a methodology for a service robot to put away the groceries in a precise and stable configuration that satisfies the desired semantic relationship of grocery and pantry objects. This task is invaluable to the motor impaired and increases the capability of contextualized service robots. We show that augmenting traditional geometric assessment with contextual knowledge of food categorization and a language association model, allows the robot to complete this task efficiently. We quantitatively validate our approach with a data set of 51 common grocery items, of which 11 objects are used for real world experiments. According to our evaluation, our method is able to successfully shelve objects next to semantically related objects 100% of the time when these relationships exist. We achieve this with an average placement precision of 3.2cm and a standard deviation of 1.1cm. We discuss remaining challenges and needed improvements for robots with these capabilities to be introduced to home environments.

3.1 Introduction

Online grocery purchases have grown 40% annually the last 5 years [44]. This service is invaluable to the elderly and mobility impaired who need significant assistance to lift and carry grocery products. However, grocery delivery services rely on the homeowner to retrieve, sort, and store their groceries after they are delivered to the front door. The elderly and mobility impaired may be unable to move their groceries from the delivery point (i.e. the front door) to

Figure 3.1. Robot asked to put away sugar box object (left) and test environment with all grocery objects (right).

their pantries. Significant attention is already being paid to how in-home service robots can help such populations with other activities of daily living[41]. Thus, it would be an important and (almost) immediately applicable service for in-home assistive robots to be able to sort and shelve groceries inside the home.

This task is challenging for a number of reasons. Every kitchen is organized slightly differently. Typically, bakery goods are in one section, beverages (soda, coffee, etc) in another section and cooking in a third. Within the pantry, the organization is also clustered. Pasta is in one section, cans in another, flour, sugar, sauces, and condiments in other sections. Leveraging knowledge of food groupings, it is possible to cluster objects into major grocery categories.

Recognizing the current layout of objects in a kitchen can facilitate an understanding of the existing organizing scheme, while also providing strong priors over where a new set of bought grocery items should be placed. This allows for per-home user customization even though organization schemes are likely to vary from one home to the next.

For a grocery item being put away, once an area/shelf has been identified, the next task is to find open space next to related items to allow for placement. Putting an object into a confined space is challenging but possible. Our work presents an approach to sort through a set of groceries using a built ”categorical” map of a kitchen pantries, and then perform the manipulation actions to put away new groceries precisely.

Our contributions are as follows: (i) we develop 4 types of contextual knowledge in the form of a hierarchical food categorization based on previous experts' food ontologies. We formalize and integrate this hierarchical categorization with a data-driven language association model that can be used to determine food groupings when putting away a grocery item. (ii) we demonstrate and evaluate our approach on a robotic system in a real world scenario. We measure success in terms of an object's placement satisfying a desired semantic relationship and geometric distance of the object placements to the desired placement.

In section 3.2 we present related work. Section 3.3 - **??** presents the methods to categorize groceries, identification of objects to map the local pantry and the manipulation strategy to put away the groceries. Section **??** presents real world experimental results and finally in Section **??** and **??** we discuss conclusions and point to open challenges.

3.2 Related Work

3.2.1 Service Robotics

In 1989, Joseph Engelberger predicted that we would soon have service robots in our homes [42]. It took a while before we saw real robots in the home. A good early example is the Care-o-Bot system developed by the Fraunhofer Institute [52]. Other classical works in service robotics include [161, 183, 77]. In [77], which demonstrate the ability of a mobile robot to approach and grasp everyday objects that are important for motor-impaired users. In [161] they show a robotic system that performs tasks of object retrieval with the addition of environment manipulation. In these efforts, the robots' sense-plan-act primitives interact with each other according to the control architecture paradigm chosen. In both hierarchical and hybrid paradigms, contextual knowledge is generally considered only during high-level reasoning and is ignored in low-level planning of manipulation behavior.

Most present day service robotic research still follows a hierarchical or hybrid paradigm and focuses on similar tasks involving grasping and moving objects from one location to another and the ability to manipulate the environment, such as opening drawers, doors, fridges, and dishwashers. These approaches use more sophisticated methods that improve reliability, allow for interaction with unknown objects, and can give users the ability to teach the robot new skills [82, 8, 177]. Recent research on robotic organization and shelf stacking, are centered around the autonomous moving of objects and focus on changing a physical environment to a specified goal state [9]. Both [6] and [97] explore shelf tidy up tasks but in [6], their focus is categorizing products based on user preference rather than precise placement on a shelf. Furthermore, as [167] states, it is unrealistic to investigate the preferences of object pairs for each user for every destination. In [97], they also incorporate user preferences while using a dataset of images containing grocery items to learn grocery categorization. These images do not provide any hierarchical categorization and only contain 25 types of grocery items to do inference on.

In contrast to prior efforts our main focus is to leverage prior knowledge such as food categorization and ontologies to organize the handling of objects. The system is organized to enable inclusion of user preferences as needed. The objective is to manipulate objects on shelf-like structures, organizing them to be next to related objects.

3.2.2 Semantic Placement

Semantic placement is the task of picking and placing an object in a stable configuration that satisfies a desired semantic relationship [125]. In addition to this, the placement distance to other semantically related objects is considered as a metric for success. This is in accordance with metrics used in rearrangement tasks, of which semantic placement is a sub-area [9]. The use of semantics in robotics has grown rapidly in the past few years[49]. Semantic localization [130, 188] and semantic grasping [110, 98] being just two fields that have seen more research of late. In this work, we focus on semantic placement of objects. While some of the concepts in other semantic work can be incorporated in this field, semantic placement is an area of work itself.

3.3 Problem Definition

Given a pantry filled with grocery items, let $\mathscr{P}$ represent the set of all known shelved pantry items and $\mathscr{K}$ represent the set of all known kitchen grocery items, not yet shelved and needed to be put away. Each pantry item $p \in \mathscr{P}$ and kitchen item $k \in \mathscr{K}$ is represented by four vectors corresponding to hierarchical categories described in Section **??**. Hence for all p there exist: $\hat{p}_{fg}$, $\hat{p}_{cg}$, $\hat{p}_{sa}$, $\hat{p}_{pu}$, where $\hat{p}_{fg}$ represents the Food Group category, $\hat{p}_{cg}$ the Container Group, $\hat{p}_{sa}$ the Stability category and $\hat{p}_{pu}$ the Purpose category. Similarly, for all k there exist: $\hat{k}_{fg}$, $\hat{k}_{cg}$, $\hat{k}_{sa}$, $\hat{k}_{pu}$. Each vector has a different length according to the number of categories belonging to that group, and is composed of an array of binary variables, where 1s denote matched categories. Our objective is to determine the object similarity score ($\mathscr{OS}$) for each k and p pair, in order to find the item most semantically related to k.

$$\mathscr{OS} = \mathscr{W}_f * \cos(\hat{k}_{fg}, \hat{p}_{fg}) + \cos(\hat{k}_{cg}, \hat{p}_{cg}) + \cos(\hat{k}_{sa}, \hat{p}_{sa}) + \cos(\hat{k}_{pu}, \hat{p}_{pu}) + \mathscr{W}_v * \mathscr{WV}(k, p) \tag{3.1}$$

In Eq. (3.3), the cosine similarity, defined as $\cos(A,B) = \frac{AB}{\|A\|\|B\|}$, is computed between each of the hierarchical category vectors in k and p. This allows two vectors to be compared based on the closeness of their angle within the embedding space. The function $\mathscr{W}\mathscr{V}$ is the direct item pair relationship score given from the cosine similarity of *Word2Vec*'s word embedding of k and p [103]. The similarity score from each component is weighed and added to get the total object similarity score between k and p. The weights $\mathscr{W}_f$ = .5 and $\mathscr{W}_v$ = .5 used for Food Group and *Word2Vec* respectively is determined by analyzing the co-linearity within our components. We describe the analysis we perform to determine these weights in **??**.

3.4 Understanding Grocery data

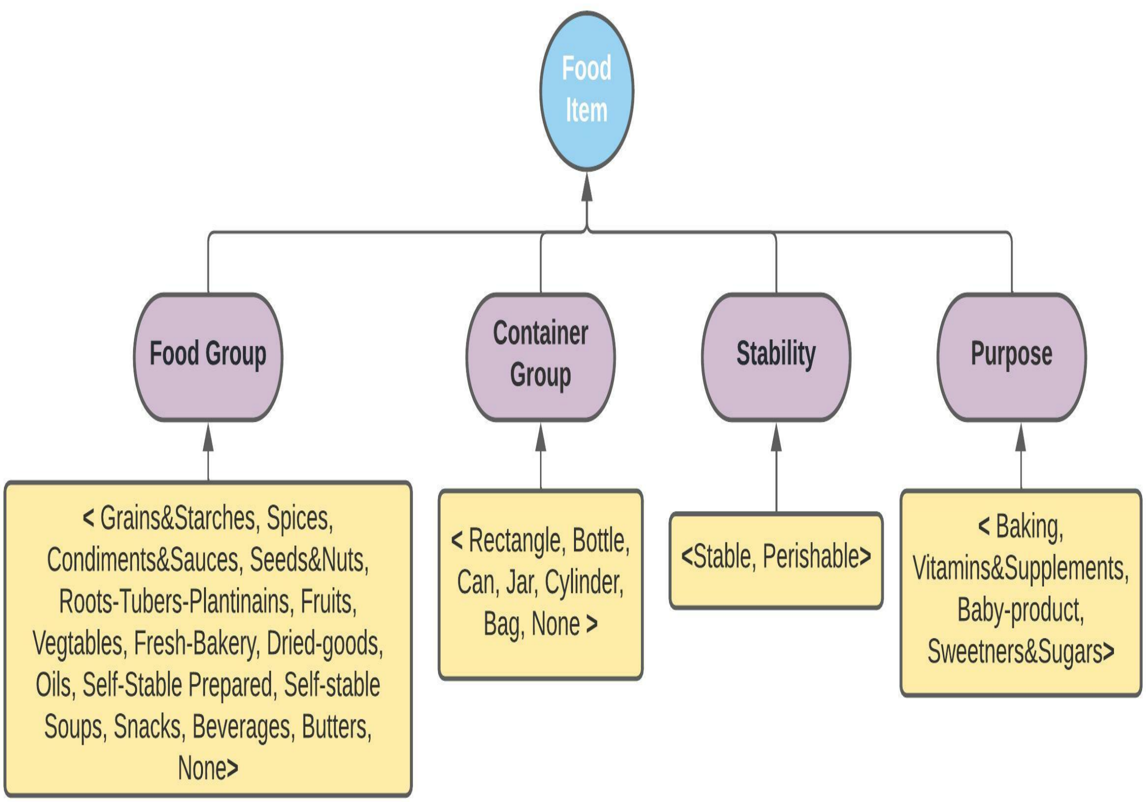

Figure 3.2. Hierarchical food categories used to represent a kitchen and pantry item where each item has an associated 4 vectors of ones and zeros. See **??** for more details.

3.4.1 Context

We create our hierarchical categories (Fig.3.2) from four primary sources:

- The Food and Agriculture Organization of the United Nations and World Health Organization ”Global Individual Food consumption data Tool” [89].
- The National Health and Nutrition Examination Survey, a branch of Centers for Disease Control and Prevention, ”What We Eat in America (WWEIA) Food Categories” [133].
- The National Institute of Health’s food groups.
- Private label and consumer packaged goods, food categories from the Canadian Grand Prix New Products Awards[2].

We chose the top categories most commonly existing in households. For instance, the categories Fruits and Vegetables were seen in all sources and descriptive categories such as Baby products and Nutrition supplements had high occurrences. We group similar categories together and thereby create a hierarchy of categories based on these groupings.

In the first level, Food Groups describe the different categories of food an item can be classified as. Container group tells the closest geometric shape of the object. For instance, a can of coke would be classified as both a Can and a Cylinder, hence, the corresponding vector $\hat{p}_{fg}$ will be [0,0,1,...1,0]. The category Stability describes if an item has an immediate shelf life of less than a week and a half, and the category Purpose describes common uses for items. This categorization gives us a way to represent a food item in a meaningful semantic manner.

In addition to the context we create above, we also want to give context of word relationships. To do this, we use Word2Vec [103], trained on the Genmi ’glove-wiki-gigaword-300’ data set. This is a language association model that we can easily incorporate as part of our context to understand the semantic relationship between two items. We show how this is done in Eq. (3.3).

Figure 3.3. Eigenvectors of our data matrix X

3.4.2 Analysis

To understand if any correlation between the five components used in our object similarity score exist, we first create a data matrix $X \in \mathscr{R}^{1326\times5}$. The 1326 rows represent all unique pairings of our 51 objects and the 5 columns correspond to our five components. Each cell is the similarity score for a set of object pairings given an individual component.

We preprocess X by standardizing according to the columns of the matrix - since the columns represent our components. We then compute the correlation matrix of X with respect to our five components. This shows the correlation that exists between our components. Food Group and Word2Vec are the most correlated components with a correlation of .53.

We further analyze the relationship between our components by plotting each component along the top three principal components axes in Principal Component Analysis (PCA)[5] (see Fig 3.3 and Fig. 3.4). We find that Food Group and Word2Vec both contribute similarly along axes one and three. This is unsurprising given their correlation.

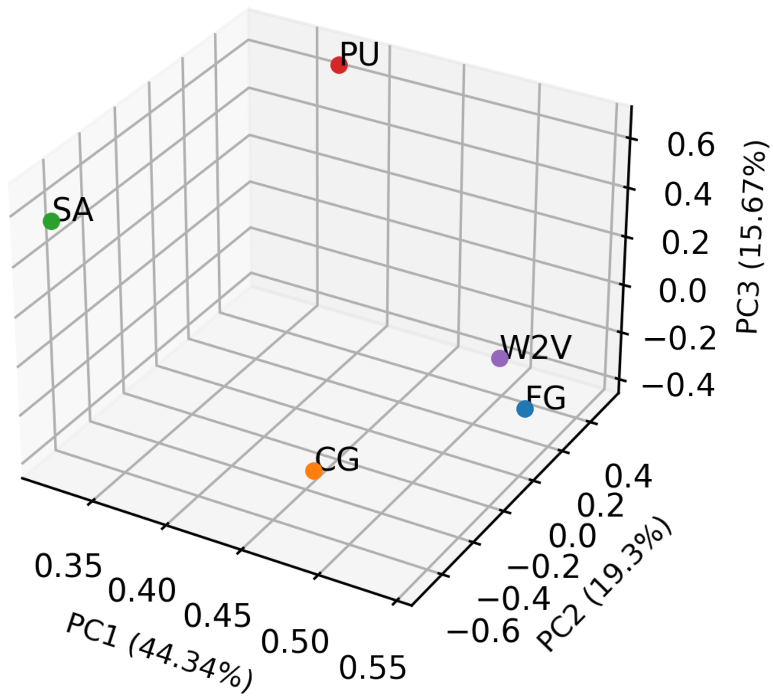

Figure 3.4. Our five components plotted along the PCA axis 1,2, and 3 of our data matrix. The contribution to the variance each PC axis has is shown next to each axis label as a percentage.

As a final comparison, we train Word2Vec on the Genmi 'glove-twitter-100' dataset as opposed to the 'glove-wiki-gigaword-300' mentioned before. We believed that training Word2Vec on a different corpus would yield different correlations results. We find this to be true as the 'glove-twitter-100' dataset leads to lower correlations between Word2Vec and Food Groups.

From these analyses, it is clear that there is some redundancy between the information provided by both Food Groups and Word2Vec given our dataset. Hence, based on their correlation, we chose to weigh both components by .5. A different training corpus for Word2Vec would give reason for different weights and possibly even the complete minimization of either Food Group or Word2Vec, given a high enough correlation.

3.5 Robotic Experiments

3.5.1 Shelf Stocking

Given a kitchen item k to be put away on the shelf, we compute the object similarity score of k to other items currently present in the pantry using Eq.3.3. The pantry item with the highest score is determined to be the most semantically related object to k. We lift the robot torso to the shelf location of the pantry item and try to identify the matching pantry item p using a trained MaskRCNN [61]. This gives the 2D bounding box over the object as well as an instance segmentation mask of the object in the image plane. If this item is not detected, we repeat the process with the next highest scoring pantry item.

Once a matching pantry item is found and detected, we need to determine a placement next to the object. To do this, we first leverage the open source ROS library rail_segmentation[3], to segment the pointcloud of objects on the shelf and shelf plane. Each segmented object label is unknown but has an associated 3D centroid point. To determine which segmented object is our recognized pantry item, we project each centroid to our image plane. The projected centroid point within our recognized item's bounding box is considered our matched pantry item's centroid. This is shown in Fig. 3.5

To determine the location for placement and check if there is free space next to our matched pantry item, we translate the centroid of the matched p to the left and right based on $k's$ known size and a threshold of 1cm based on our known $p's$ size. This translated centroid is used as the placement location if it is in free space. In rail_segmentation, a hole in the segmented shelf exists due to an segmented object at that location. Hence, if there is no hole on the segmented shelf at the location of the centroid translated and projected onto the shelf plane, that location is free of obstacles. Once we have identified a location for placement, we use an Inverse Kinematics solver (IK)[11] and MoveIt![25] to place k at this location.

Given the location for placement, we need to plan an arm motion trajectory to place the object. Using an IK solver, we first search for a feasible placing arm configuration where the

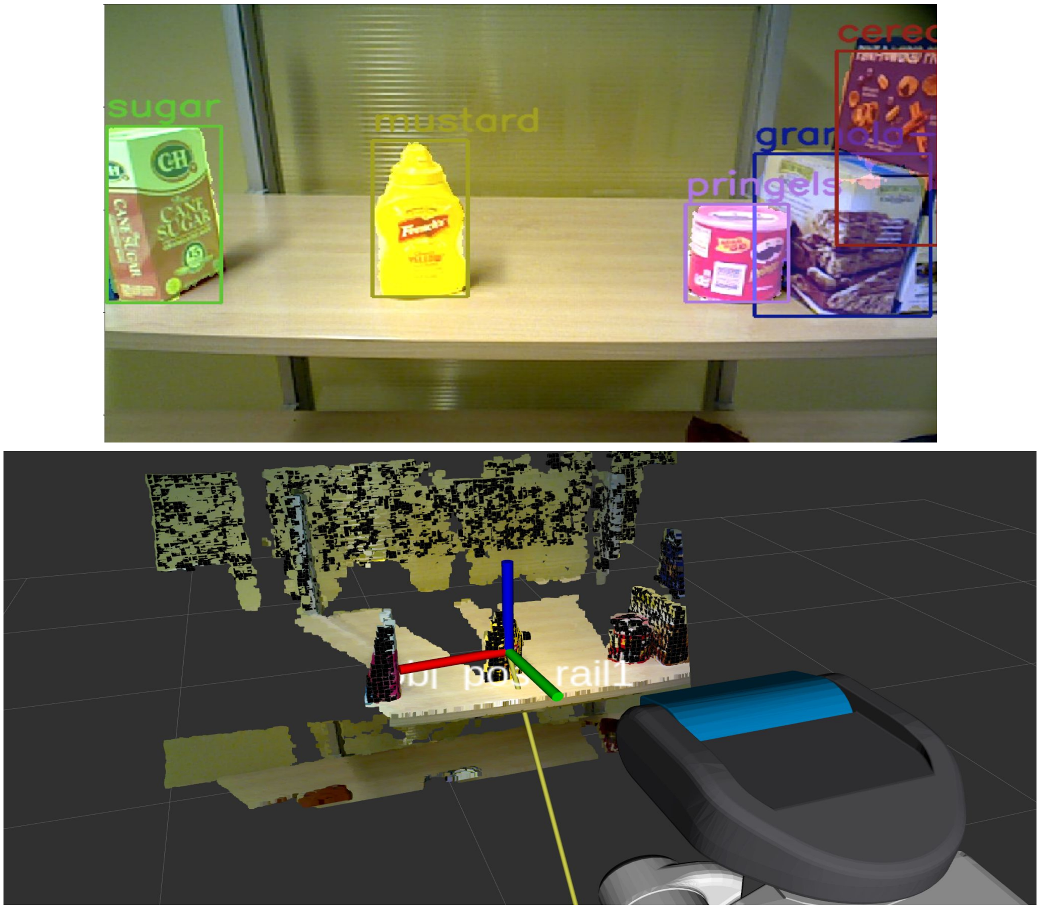

Figure 3.5. Object segmentation results from training using MaskRCNN (top). Object pointcloud segmentation results from rail_segmentation (bottom) . The black points are segmented objects and the centroid of the mustard object is returned after matching.

gripper at the placement location with its orientation constrained to have its z-axis parallel to the normal vector of the shelf. Once a feasible arm configuration is found, we plan an arm motion trajectory from the current arm configuration to the placing arm configuration by using the RRT planner available in MoveIt!. If a motion plan is not found, or the IK solver fails, we consider this a failure due to the tight constraints for planning.

3.5.2 Experiments

We use the Fetch robot [180] to perform the task of putting away four kitchen items in a two-level shelf filled with seven pantry items in our known dataset and other unknown food items. Unknown means the items are not in our dataset and not trained on our object detector.

We consider two metrics used within the rearrangement and semantic placement literature when testing our approach. First, a geometric goal where a threshold of less than 5cm from the object k to the target object p is considered "near" and therefore a precise placement. In this regard, any placement greater than 5cm is not considered successful. Second, a semantic goal where we compare the food grouping of our approach to that of user preference groupings determined from a collaborative filtering approach.

Our total labeled data set accounts for 51 common grocery items, of which 10 unique items are used for testing (Pringles, Mustard, Spam, Sugar box, Ketchup, Hot-Sauce, Granola-bar box, Cookies, Muffin-Mix box, Beans can) including one item twice (Pringles). Of these items, Pringles, Mustard, Spam, and Sugar box are from the YCB dataset[20] in order to allow for greater reproducibility and comparison.

For our first metric, we conduct 5 total placement trials. In one trial, we attempt to put away all four kitchen items in an order chosen at random. This means that the level of difficulty in the constrained motion planning task increases as the shelves become more cluttered. This is representative of a real environment where all kitchen items should be put away consecutively and in a timely manner. The four kitchen items to be put away are Pringles, Mustard, Spam and Sugar box. For each trial, the initial shelf configuration is random as the remaining seven known objects and varying unknown food objects are placed upon the shelf.

Using our clustering method desired in Section 3.3, for the seven known objects in the pantry shelf and four kitchen items, Pringles is determined to be placed next to Pringles, Mustard next to Ketchup, Sugar box next to the baking Muffin-Mix, and Spam next to Beans. We record the placement distance for each of these placements in Table 3.1 for trials 1-4. In trial 5, we remove the pantry item with the determined closest semantic meaning to our kitchen items and cluster our items again. In this new clustering, Pringles is determined to be placed next to Cookies, Mustard to Hot-sauce, Spam to Cookies, and Sugar box to Granola-bar box. The placement distance of these placements are shown in Table 3.1.

Our results show an average placement distance of 3.2cm to the desired semantic object with a standard deviation of 1.113. In all trials, the robot was able to successfully detect the desired semantic object using the MaskRCNN 100% of the time. In this way, the robot always placed a kitchen item to the pantry item with the highest semantic similarity score. Part of this success has to do with our small test set which allows for highly accurate recognition.

To test the semantic grouping of our approach, we compare the pairwise grouping of food items using our method, to users' preference groupings determined through a collaborative filtering approach [6]. Compared to our approach, which uses established expert features, as well as a word association model, for describing and understanding food grouping, this approach relies on user input to determine prefered groupings. Due to limited information for reconstructing user preferences and test scenarios, we compare with a small set of items. Given the same shelf configuration of food items and six food items to be put away, our approach groups the following items similarly: Rice next to Pasta, Cake Mix next to Flour, Coffee next to Tea, and Pasta next to Rice. For two shelf configurations, our approach differs when grouping Sugar and Cake Mix. Our approach groups Sugar with Tea and Cake Mix with Flour, while the user preference approach groups Sugar with Salt and determines Cake Mix should be stored separately. Because of limited access to user preferences data, it is unclear if these cases were an error in determining the user preference or if these groupings are truly based on the users preference given the shelf configuration. Nonetheless, we find that for four out of the six cases, our groupings largely match those determined by users' preferences groupings, even without input from the user. To allow our work to be easily compared to future work, we have listed all 10 unique items used in our experiment and plan to release our code for food groupings at publishing time (https://github.com/cogrob/putting_away_groceries).

In Section 3.5.3 We discuss some of the failures that occurred during our robot experiment trials, other potential areas for improvement, as well as, give more insights on the meaning of trial 5

Table 3.1. The performance of placements where each entry tells how close the grocery object was placed to the target pantry item. ”X” indicates a failure. We discuss these failures in section 3.5.2.

Grocery Objects	Trial 1 (cm)	Trial 2 (cm)	Trial 3 (cm)	Trial 4 (cm)	Trail 5 (cm)	Mean (cm)	STD (cm)
Pringles	4.4	2.3	2.5	3.6	X	3.2	0.851
Spam	.8	2.9	3.5	X	3.2	2.6	1.06
Mustard	3.6	X (5.7)	3.9	3.2	X	4.1	0.956
Sugar	X	4.5	2	2.8	2.3	2.9	.0966
Total Error						**3.2**	**1.113**

3.5.3 Discussion

Each ”X” in Table 3.1 represents a failure during the trial. These failures fall within two primary categories: rail_segmentation failure and motion planning failure.

While testing our approach, we noticed that despite detecting the correct semantic pantry item through our MaskRCNN, no centroid from the segmented objects on the shelf were found within our detected item’s bounding box. Hence, no matching segmented object was found. Through further testing, we found that the centroid given from rail_segmentation was not always the centroid of all segmented objects as we expected. While open space segmentation worked well, in high clutter separate objects were sometimes considered one object, changing the location of the expected centroid. An approach to solving this problem would be to directly perform pose estimation of our object. In this way, we can ensure the 3D centroid we generate belongs to our target object.

The second cause for failure we saw while testing resulted from motion planning failures. Some shelf configurations were too constrained for our motion planner to find a suitable arm trajectory to the desired placement arm configuration. A possible solution to this in later work would involve repositioning items to make more space.

In spite of the few failures that occur, we see that our approach is able to find meaningful semantic relationships when these relationships are present within our kitchen and pantry items. This is especially highlighted through the placements in trial 5 where the semantic similarity score for Pringles, Mustard, and Sugar box is high in comparison to Spam.

3.6 Conclusions

In this work, we present a detailed and empirically validated method to have a robotic system sort and shelf grocery items into a pantry - satisfying desired semantic relationships. We introduce and formulate our approach to determine the semantic relationships of grocery item pairs. Our method shows consistent statistically valid placements according to the geometric placement metrics used within the rearrangement community. This task is pertinent for elderly and mobility impaired, for whom lifting and carrying grocery products can be infeasible and aligns well with modern grocery delivery services.

3.7 Acknowledgements

I thank Jiaming Hu for his assistance with pose estimation of objects. This chapter contains material from ”Putting away the groceries with precise semantic placements.” by Akanimoh Adeleye, Jiaming Hu, and Henrik I Christensen, which appears in Proceedings of IEEE 18th International Conference on Automation Science and Engineering (CASE), 2022. The dissertation author was the primary investigator and first author of this work.

Chapter 4

Global Home organization of Objects

We consider the problem of building an assistant robot that can help humans in daily household cleanup tasks. The creation of such an autonomous system in real-world environments is inherently a challenging multi-objective task, comprising – (i) Detection of objects out-of-place and prediction of potentially correct placements, (ii) Fine-grained manipulation for stable object grasping, and (iii) Room-to-room navigation for transferring objects in unseen environments. This work systematically tackles each component and integrates them into a complete object rearrangement pipeline. To validate our proposed system, we conduct multiple experiments on a real robotic platform involving both long-horizon and complex pick-and-place tasks. Additional details regarding the code and video demonstrations will be available at https://sites.google.com/eng.ucsd.edu/home-robot.

4.1 Introduction

Creating autonomous agents that can aid human beings in everyday household chores has long been considered the holy grail of service robotics research. In this work, we take a step towards that goal by proposing a complete system for identifying misplaced objects in the environment and transferring them to their desired locations. There are several aspects to this inherently long-horizon task that make it particularly challenging in a real-world environment. Firstly, recognizing objects out of place in a noisy environment is a non-trivial problem. While

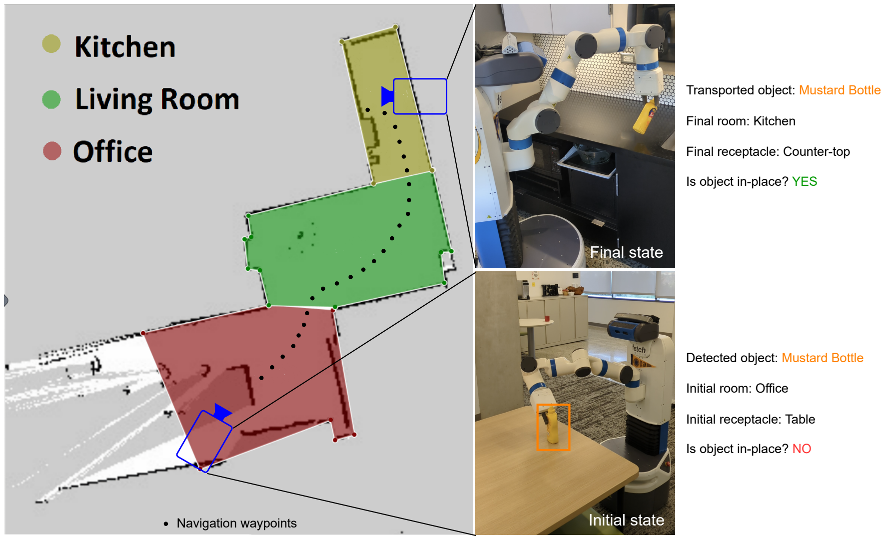

Figure 4.1. The overview of a home-robot object rearrangement task. At the initial state, the robot identifies the `mustard_bottle` object, and determines that it is out-of-place in the `office`. Subsequently, it picks up the object using stable manipulation technique, and transports it to its correct location in the `kitchen` on top of the `counter-top`. The semantic map used for the navigation task is shown on the left with the robot's trajectory overlaid on it.

state-of-the-art open-vocabulary object detectors [186, 56, 106, 189] are quite adept at localizing objects in a zero-shot manner, determining whether they belong in a particular environment is more complicated, as it also involves understanding scene-context and learning about user preference of object placements on target surfaces (hereafter called *receptacles*). Secondly, stable grasping and manipulation of objects in cluttered environment is still an open research problem due to the difficulty of affordance estimation and motion planning. Finally, transferring the held object to a different room that has previously not been annotated *completely*, is difficult because the accurate locations of objects in the target room, including the target receptacle surface where object needs to be placed, are considered unknown.

In this paper, we systematically tackle each individual component and integrate them into a complete pipeline for rearranging house-hold objects in a real-world environment, using a physical mobile manipulation system. Figure 4.1 shows an example of a home-robot rearrangement task that we consider in this work. At the initial state, the robot detects an object `mustard_bottle`, and determines that it is out-of-place in the `office` environment atop a `table`.

Using a reasoning module, it further predicts a probable correct room and receptacle surface location for it. In this case, the predicted room is `kitchen`, and the receptacle is `counter-top`. Then, the robot picks up the misplaced object by stable grasping, and plans a path to the target room using a constructed semantic map of the environment. Upon reaching the `kitchen`, the robot scans the surroundings for a possible `counter-top`. Once it has succeeded in correctly localizing it, the robot plans a path to reach the `counter-top`, and places the held `mustard_bottle` safely on the receptacle.

The rest of the paper is organized as follows. Section 4.2 discusses existing approaches for object rearrangement in home-robot environments. Section 4.3 has the details of each of the components we use to perform the overall task, with a summary of the integrated system in Section 4.4. We explain the details of experiments conducted in Section 4.5, and provide a summary of our work with some future goals in Section 4.6.

4.2 Related Work

Recently, indoor object rearrangement tasks using mobile robots have gained much popularity. Due to the increasing number of Embodied AI platforms available [128, 154, 155, 166, 13, 96], several methods have been proposed for solving the complete mobile manipulation task end-to-end. However, these approaches [71, 105, 128, 9, 166] are mostly trained in simulation, and rarely generalize to real-world environments. Other works have adopted the task planning approach but are either restricted to specific tasks such as folding clothes [162] and rearranging kitchens [181], or follow a pre-defined template [30]. Some approaches [149, 53, 132, 84] focus on the human-robot interaction aspect but not on autonomy. Two efforts closest to ours are that of Wu *et al.*[182] and Castro *et al.*[21]. Wu *et al.*[182] use large-language models (LLMs) to infer generalized user preferences and use it to tidy a room. However, they do not handle fine-grained manipulation, need rigorous prompt engineering to understand user preferences, and are limited to within-room navigation. While Castro *et al.*[21] do consider room-to-room navigation, they rely on manually annotated prior semantic maps for querying the exact locations of rooms and objects. In contrast, we only build a simple 2D geometric map with rough room locations and proceed with identifying target locations of target receptacles in the environment for object placement.

4.3 Components

In this section, we will go over the individual components that together make up our ensemble system for rearranging objects in a home environment.

4.3.1 Semantic mapping and visual recognition

The first step of our pipeline involved construction of a map of the environment. We use the Cartographer system [64] for generating a LiDAR-based occupancy-grid map of our environment. This gives us a 2D representation of our world in terms of obstacle, and free space. Subsequently, using this geometric map, and RGB images obtained from the first-person view camera of the robot, we manually annotated the locations in a scene with a semantic label of the room category.

To recognize entities in the surrounding, we use a recently proposed DETIC [189] model, which has been trained to detect twenty-thousand object classes using image-level annotations. We use the detector in two-stages – (i) To detect target objects that we can manipulate, we extract 2D bounding-boxes from the images for precise localization, and (ii) For detecting receptacle surfaces which are often irregularly shaped, we extract segmentation masks which roughly identifies the region corresponding to the receptacles.

4.3.2 Object rearrangement

In this work, our motivation for rearranging homes is to do so based on both *common-sense reasoning* (for target *rooms*), and *human preference* (for target *receptacles*). We leverage a recently proposed large dataset of human-labeled preferences for object placement location in

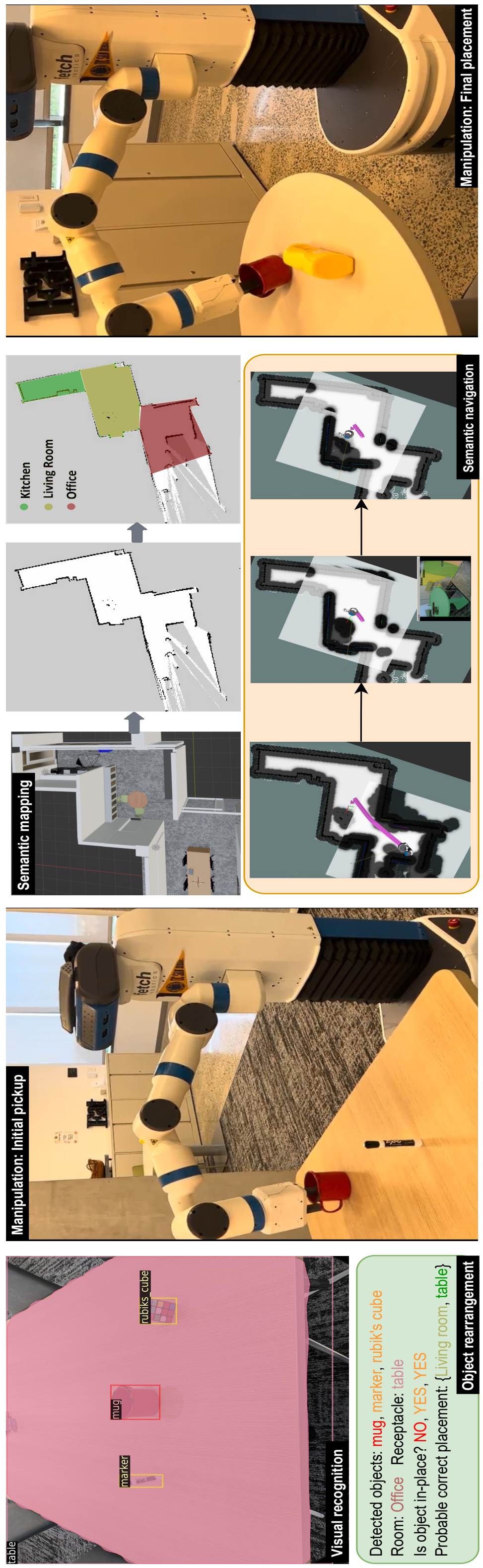

Figure 4. All components of the proposed system

a home [79] environment as our object vocabulary. Using this, we create a knowledge base of most likely room locations for a particular object based on a majority vote. This corresponds to common-sense reasoning of which rooms are likely candidates for hosting a particular object. For predicting target receptacle locations, we utilize user preference to learn diversity in human choices. However, as the dataset does not contain a particular user's preference for *all* the objects, we adopted a recommender system approach for matrix completion. We build on existing work that tries to use user prefaces through collaborative filtering to put away groceries and toys [6].Collaborative filtering is a recommender system where, given scores for users ranked preferences, we can construct a sparse user matrix and perform matrix factorization [86]. This allows us to predict user ratings $r_{u,i}$ for user u and item i. In our case, an item i refers to the objects placement in a room and on a receptacle. We can predict a user's rating using the equation $f(u,i) = \gamma_u \cdot \gamma_i$. Here, $\gamma_u \in R^d$ and $\gamma_i \in \mathbb{R}^d$ are latent vectors representing the row of a user in matrix γ_U and the column of an item in matrix γ_I; where d is the lower dimensional space.

To chose our parameters $\gamma = \{\gamma_u, \gamma_i\}$ to most closely fit the data, we minimize a loss function using Mean Squared Error plus an L2 regularization term.

$$\arg\min_{\gamma} \frac{1}{|\tau|} \sum_{r_{u,i} \in \tau} w_{u,i}(r_{u,i} - f(u,i))^2 + \lambda \Omega(\gamma) \tag{4.1}$$

where τ is our corpus of ratings and $\Omega(\gamma)$ is $\ell 2$ norm $||\gamma||_2^2$. Our completed approach allows us to estimate the full preference of users' desired correct object placement locations.

The overall task of object rearrangement involves primarily two steps: (i) Out-of-place object identification – For this, we enumerate a list of top-k ($k = 10$ in this work) likely locations (room and receptacle) based on our completed user matrix. If the current room and receptacle location for a particular object does not fall in this list, we classify it as out-of-place. (ii) Preference-based object placement prediction – For every out-of-place object, we first use common-sense reasoning to determine its target room location. Subsequently, the various preference-based target receptacle locations corresponding to that room can be obtained based

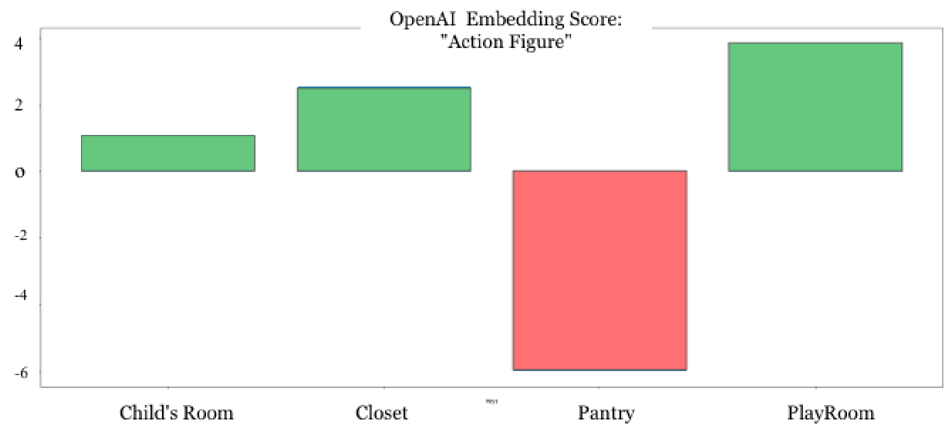

Figure 4.3. Example scaled similarity score calculated from the LLM word embedding of `text-davinci-003`, a variant of GPT-3

on a sampled user identity.

4.3.3 Large Language models in Collaborative Filtering

In Equation 4.1, we define and show our loss function for matrix factorization as part of our collaborative filtering method. One thought we explore in this project, is the idea of using the general semantic relational knowledge and understanding of LLMs to help guide our function toward reasonable scores an average set of users might give.

To do this, we augment Equation 4.1 as following:

$$\arg\min_{\gamma} \frac{1}{|\tau|} \sum_{r_{u,i} \in \tau} w_{u,i}(r_{u,i} - f(u,i))^2 + \lambda_1 \Omega(\gamma) + \lambda_2(\overline{UP}_{u,i} - OP_{u,i})^2 \tag{4.2}$$

In this equation, *OP* represents our incorporated LLM, `text-davinci-003` - a variant of GPT-3 [17], by using the models word embedding space. Given two semantic labels, the similarity between the two labels are calculated from the word embedding using the cosine distance. This score is then scaled to match the range of our user data. An exemplary graph of this is shown in Fig 4.3. In this example the scores for the semantic relations between `Action figure` and `Child’s room`, `Closet`, `Pantry room` and `Playroom` are shown. Although not always as expected, there is semantic meaning in these relations that meaning aligns with our intuition.

Because, the semantic relational scores from our LLM is not specific to individual users, but rather general knowledge about object-room-locations, When incorporating the scores of our LLM, we compare to the average scores of our users object placements to the scores of our LLM. Here $\overline{UP} = \frac{1}{|\tau|}\sum_0^u r_{u,i}$, representing the average score of all our users across all items i.

Our goal with this modified loss function is to encourage the average placement scores of our user matrix approximation to match our LLM's score.

4.3.4 Manipulation of objects

The manipulation module in this project comprises three essential components. Firstly, planning involves understanding the environment and constructing a comprehensive planning scene. Secondly, it requires to carefully analyzing potential methods of interaction with the target object. Lastly, it needs to meticulously plan the motion required to effectively interact with the object, all while keeping the task goal in mind.

Before constructing the planning scene, the robot in this project possesses some prior knowledge of the environment. For instance, it understands that most objects should be positioned on the table. Therefore, the table serves as a common obstacle during our manipulation tasks, making it beneficial to prioritize its search once the object is detected. As a large connected object, the table is added as a single entity to the planning scene, optimizing resource usage for collision detection. Conversely, for the remaining non-target objects, a voxel set efficiently represents them [26].

Even though the robot knows the planning scene, interacting with the target object is crucial. In this work, grasping is the prevailing contact approach. Hence, a DL-based grasping predictor [165] is utilized to estimate a set of possible grasping poses. However, pick-and-place is not the only manipulation action available. The robot must also account for potential object motions based on the task requirements. For instance, it might need to open a cabinet before placing an object. Consequently, the robot must compute the required motion to open it after identifying a set of arm configurations to grasp the cabinet handle. The robot may explore alternative approaches if the motion is found before the timeout.

When planning the actual motion, the robot must consider various constraints that may arise during the execution, such as delivering a cup of water or opening a cabinet. Hence, the project employs CBIRRT (Constrained Bi-directional Rapidly-exploring Random Trees) [12] to address these constraints effectively.

4.3.5 Navigation

The overall movement of the robot in our home environment is considered in two stages – (i) room-to-room navigation, for planning the path from a source room to the target room, and (ii) receptacle navigation, for navigating to the correct receptacle in the target room for object placement. For room-to-room navigation, the 2D coordinate of the center of the target room is first computed from the annotated semantic map. Using this destination point, a heuristic point-goal navigation algorithm is adopted to plan a trajectory by avoiding obstacles along the way with the Navfn planner. Once the room is reached, the receptacle navigation module is fired. First, the entire room is scanned for possible receptacles for the held object, and the position of each candidate receptacle is updated in the map by reprojecting the detected object from the depth map of the fetch camera. Then, the most likely target receptacle is chosen using the out-of-place module 4.3.2. Finally, a second heuristic planner is called to make the robot move as close to the goal receptacle position as is feasible in collision-free space, which is achieved through the Carrot Planner.

4.4 Systems Integration

4.4.1 Task Setup and Flow

Figure 4.4 depicts a block diagram of overall architecture of the proposed system. At the start of an experimental trial, objects and their initial misplacements are picked based on the

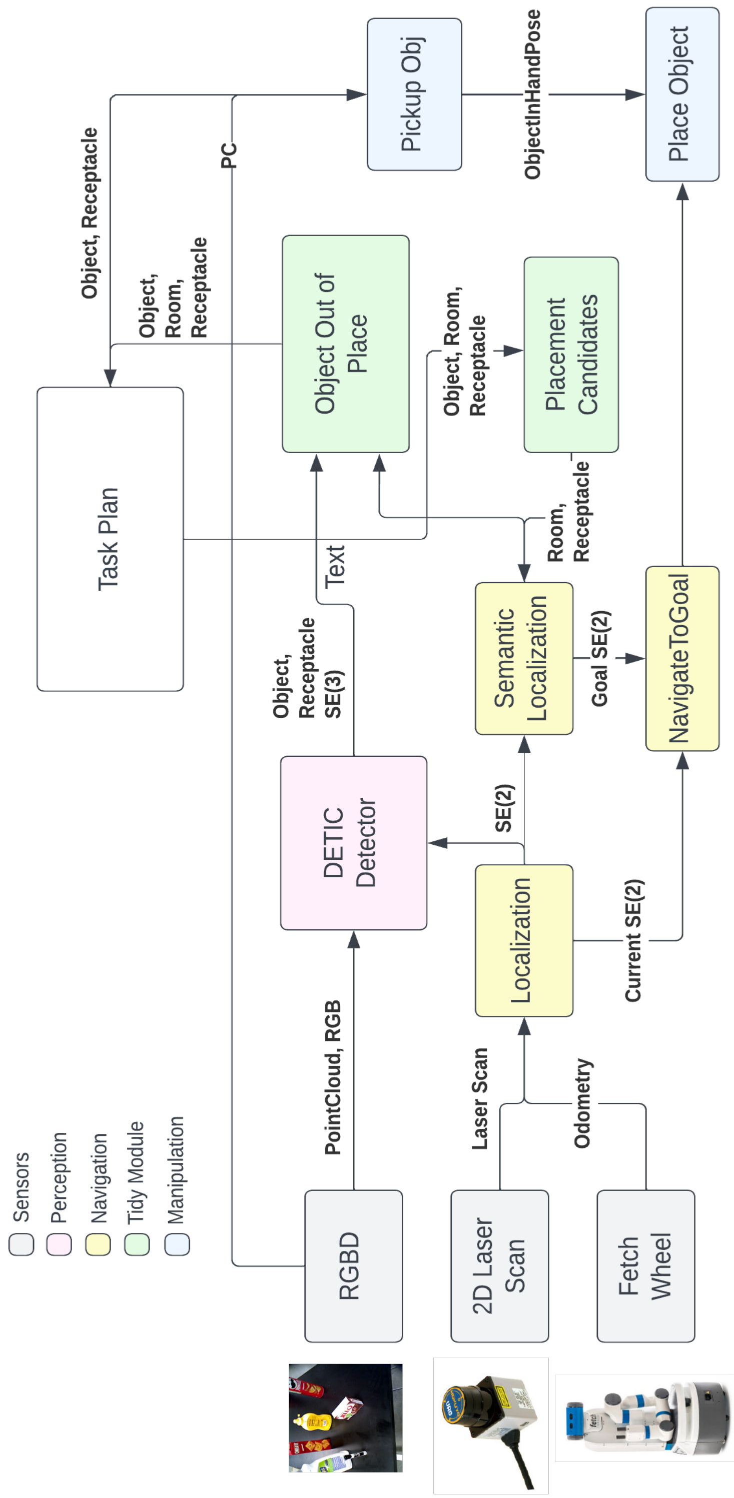

Figure 4.4. The overall architecture of the proposed system

experiment. The robot's initial geometric location within the environment's 2D occupancy-grid map is provided and later used to identify our semantic room location. The goal is to have all the objects in the environment in their "correct" locations. Given the initial room, the robot identifies objects and receptacles within its field of view. Once an object is found on an incorrect receptacle, it is added to a list of misplaced objects. The robot then prioritizes picking up the object and placing it in the correct location, given the users' preferences. The Object Rearrangement module provides the placement candidates arranged room-wise. The robot begins navigating to the target room and scans the room for receptacles. Once a suitable receptacle is found, it navigates to the goal receptacles and places the object. If a suitable receptacle is not found, then the next candidate room is tried until the object is placed. The process continues until the robot either makes an unrecoverable mistake or all items are correctly placed.

4.4.2 Use of Behavior Trees for Integration

A key component of the complex home-robot system is the composition of the different capabilities of the robot to execute the task robustly and continuously. This calls for a control architecture that is modular and capable of switching between tasks such that the different tasks can be called anywhere during the workflow. Consequently, Behavior Trees (BTs) are used to monitor and orchestrate the flow of the entire system. BTs is a modular control architecture developed for controlling autonomous agents that supports reactive behavior. [27] A BT consists of control nodes and leaf nodes, where the leaf nodes are atomic operations that include actuation and sensing. In contrast, the control nodes are behavior nodes that chain together multiple nodes. The main control nodes are sequence, selector (or fallback) and decorator nodes. Each node (with its children) is a behavior that the robot can exhibit. A behavior can be composed of multiple behaviors. For instance, picking up a misplaced object is a behavior that is composed of two behaviors: identifying a misplaced object and picking up a target object.

Figure 3 shows the BT of the home-robot tidy module. The entire system runs continuously, where the misplaced object identification (OOP) module 4.3.2 is called. Then for every

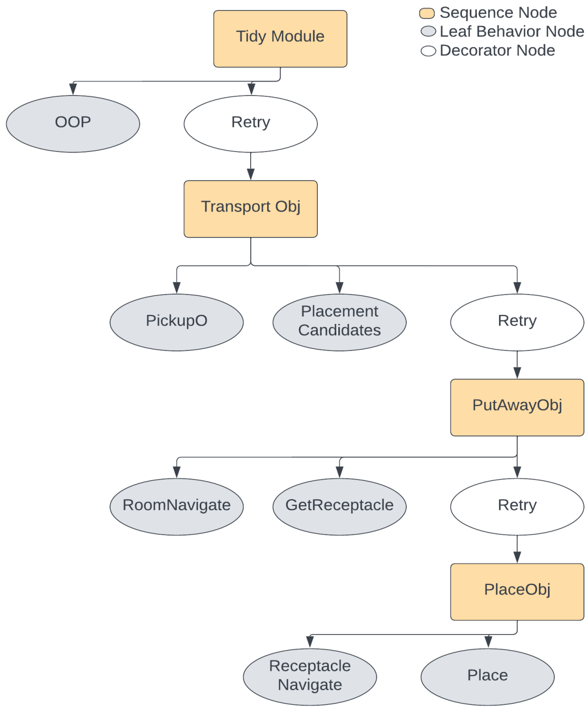

Figure 3. The complete behavior tree of the home-robot tidy module

object, potential placement candidates (PlacementCandidates) are computed 4.3.2. The Pickup Behavior 4.3.4 is called on the misplaced object. RoomNavigator, followed by ReceptacleNavigator modules are executed, given by the placement candidates computed. The PlaceBehavior is finally called to place the object. If the place action fails, then the robot tries other candidate receptacles until one succeeds, highlighting the BT's advantages. This is implemented through multiple Decorator Nodes that can facilitate retry behaviors. The different messages from each behavior are passed around through blackboard mechanisms.

For this work, the `py_trees`[1] library is used, due to its pythonic nature, and ROS support.

4.4.3 The ROS integration

The entire system is written within the ROS framework. Each module is written as an independent ROS node, and exposed either as a ROS service or a ROS action node. As per the

[1] https://github.com/splintered-reality/py_trees

Behavior Tree, sensing and computation nodes are ROS Services while actuation nodes (such as navigate and manipulate) are ROS Actions. The perception node (for objects and receptacle detection) executes continuously in the background and publishes detections at a constant rate, that the different nodes subscribe to.

4.5 Experiments

We conduct three fundamental experiments to elucidate the effectiveness of our comprehensive system. The first is a tidy-up task with two trials of the same experimental setup of misplaced items but using two different user preferences for each trial. The second is what we consider a long-horizon tidy-up task, where the robot must visit and identify at least two misplaced items in two different rooms. Lastly, we perform a complex interaction experiment where the robot must return a misplaced item to a closed environment receptacle such as a cabinet or fridge. This requires the robot to identify the environmental receptacle, then open and close the receptacle before and after placement. All experiments are performed in the real world using a mock apartment environment, created from an actual communal office space within our Lab. This space consists of a living room, kitchen, and office space and has a total of 14 receptacles.

4.5.1 User Preferences

Determining users placements

In section 4.3.2 we propose two loss fuctions for our collobarative filtering methods. Equation 4.1 uses the traditonal MSE loss with an L2 regularization term and Equation 4.2 augments this loss to incorporate knowledge from an LLM. We evaluate these two loss functions along with the base MSE loss function with no L2 regularize.

For our experiments we run 10 cross validation for each loss function by splitting our train and test set randomly 10 times and then taking the average score. Fig 4.6 shows the moving average for our three loss functions. Our results show that, using only 2% of our users data, we are able to estimate their preferences about 20 to 30% of the time. Surprisingly, our LLM term,

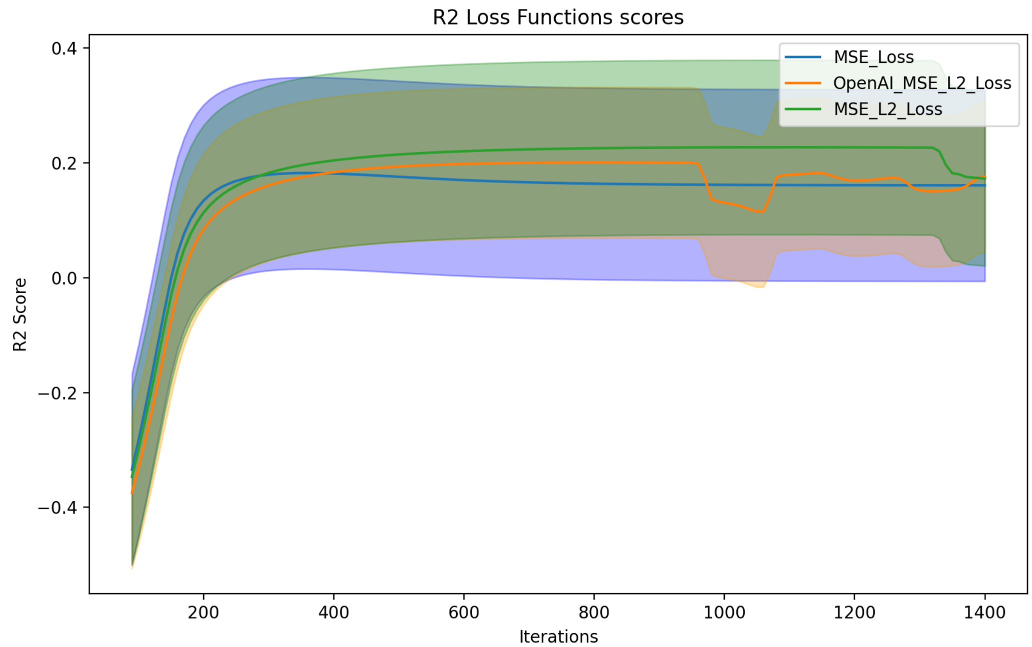

Figure 4.6. Moving average R2 scores from our three loss functions.

did not help us as much as we expected. We believe this mostly due to discrepancies between our LLM and individual user preferences. This highlights the complexity of this problem as even when using large data sets and general knowledge, it is hard to come to a consensus of where objects should be placed in the individual users homes.

User placements

An example of different user preferences is shown in Table 4.1. For all six of our real-world objects, we show the resulting placements receptacles for two different users based on our module. In this experimental setup up we conduct two trials, selecting users with different object placement preferences for each trial. We use the mug as an example object in our trails. User A prefers the mug to be placed on the kitchen counter-top or kitchen sink, while User B prefers the mug to be placed on the living room table or office table. According to our model, both users consider their placements as correct placement locations and rank them higher or equal to other possible correct placements. At the start of each experiment trial, the mugs are

Table 4.1. Example table with user placement preferences

Objects	User A		User B	
	Preferred rooms	Preferred receptacles	Preferred rooms	Preferred receptacles
rubik's cube	office kitchen livingroom	[shelf, table] [counter, table] [cabinet, table]	livingroom office kitchen	[cabinet, table] [table, cabinet] [cabinet, table]
mustard bottle	kitchen livingroom office	[cabinet, counter] [table, sofa] [table, cabinet]	kitchen livingroom office	[shelf, counter] [table, cabinet] [cabinet, table]
marker	livingroom office kitchen	[cabinet, shelf] [table, cabinet] [cabinet, table]	office kitchen livingroom	[table, cabinet] [table, cabinet] [table, shelf]
cracker box	kitchen livingroom office	[cabinet, table] [cabinet, table] [cabinet, shelf]	office kitchen livingroom	[shelf, cabinet] [cabinet, table] [cabinet, sofa]
bleach cleanser	livingroom office kitchen	[cabinet, table] [shelf, table] [shelf, cabinet]	office kitchen livingroom	[shelf, table] [cabinet, table] [table, cabinet]
gelatin box	office kitchen livingroom	[table, shelf] [cabinet, counter] [cabinet, table]	livingroom office kitchen	[table, cabinet] [table, shelf] [cabinet, counter]
potted meat can	kitchen livingroom office	[counter, shelf] [cabinet, table] [cabinet, table]	office kitchen livingroom	[cabinet, table] [counter, shelf] [cabinet, table]
mug	kitchen livingroom office	[counter, sink] [shelf, sofa] [cabinet, table]	livingroom office kitchen	[table, shelf] [cabinet, table] [sink, cabinet]
soup can	livingroom kitchen office	[table, cabinet] [cabinet, counter] [cabinet, table]	office kitchen livingroom	[cabinet, shelf] [cabinet, shelf] [sofa, cabinet]

placed at an incorrect location for both users. The robot must now tidy the environment and place the mug and other objects in the correct location given User A's or User B's preferences.

4.5.2 Long Horizon

The setup for our long horizon experiment is akin to our user preference study, except we select only one user's preference, and all objects are then placed either correctly or incorrectly at random. We use 3 to 5 objects for this experiment and have at least two placed incorrectly. Once the trial begins, the robot must identify objects and their receptacles to determine if the object is misplaced. All misplaced objects must be moved to their correct location based on the user's preference. We leave the robot to explore and move objects until at least two misplaced items in two different rooms are discovered and correctly rearranged.

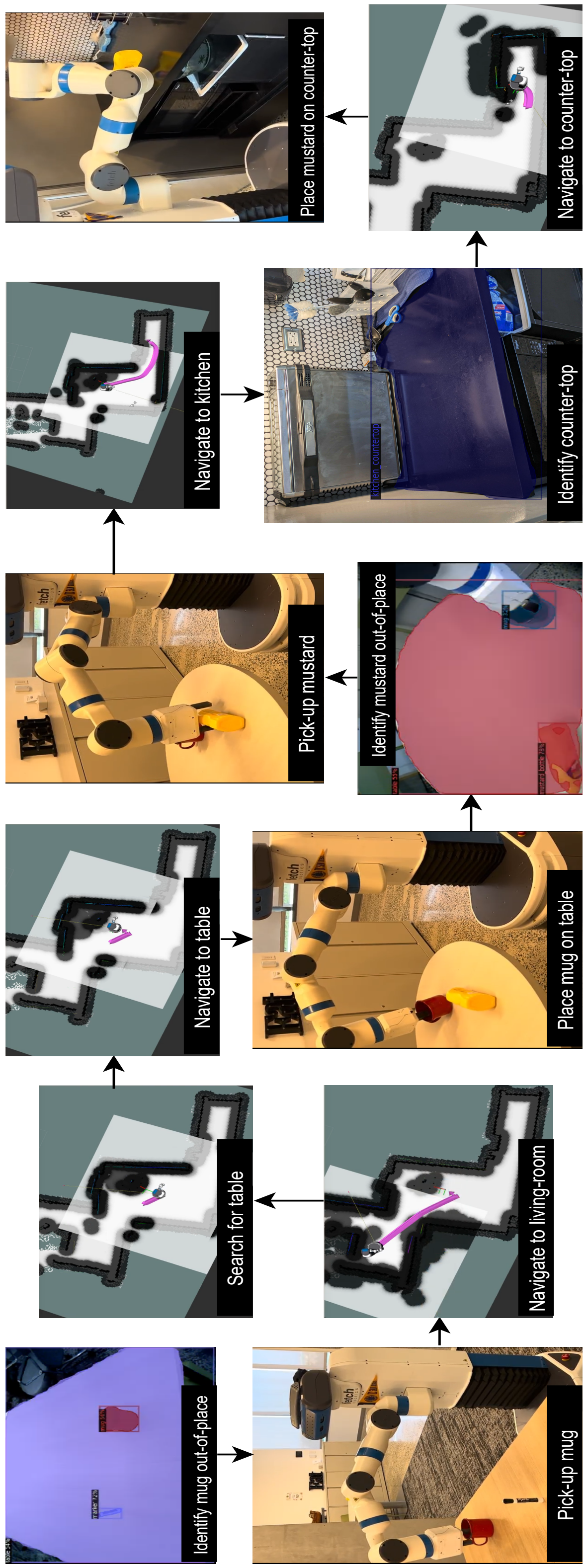

Figure 4.7. Long Horizon rearrangement

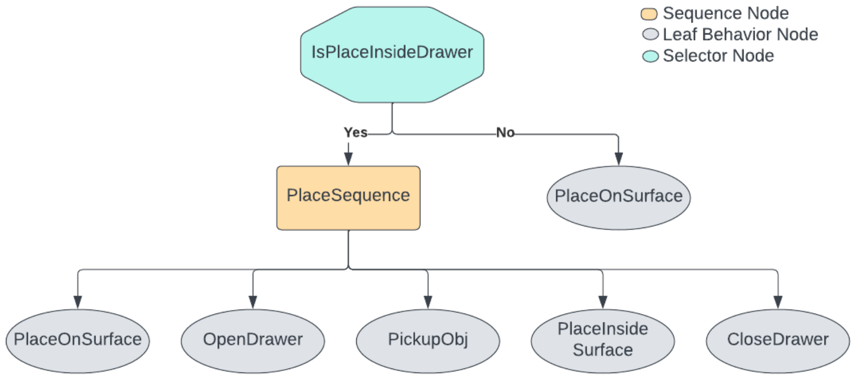

Figure 4.8. Complex Manipulation Interactions of Placing inside Drawer

4.5.3 Complex interactions

During a rearrangement task, simply picking and placing objects should not be the only manipulation the robot requires. Specifically, there may be instances where the desired manipulation is not feasible due to the state of the environment, making it crucial for the robot to have the ability to interact with its surroundings for subsequent manipulations. In our project, we illustrate this concept by using the example of placing an object, such as a Rubik's Cube, into a cabinet. The process may be more complex if the cabinet is closed, requiring the robot to perform multiple sub-tasks, as shown in the behavior tree 4.8. First, the robot must estimate a temporary location for the Rubik's Cube and the appropriate grasp poses to open the cabinet. Next, it needs to select an optimal position that facilitates placing the cube and opening the cabinet. Finally, the robot will place the cube on a surface, open the cabinet, and then pick up and place the cube into the cabinet. This example highlights the importance of flexibility and adaptability in robotic manipulation, ensuring the robot can interact successfully with a dynamic environment.

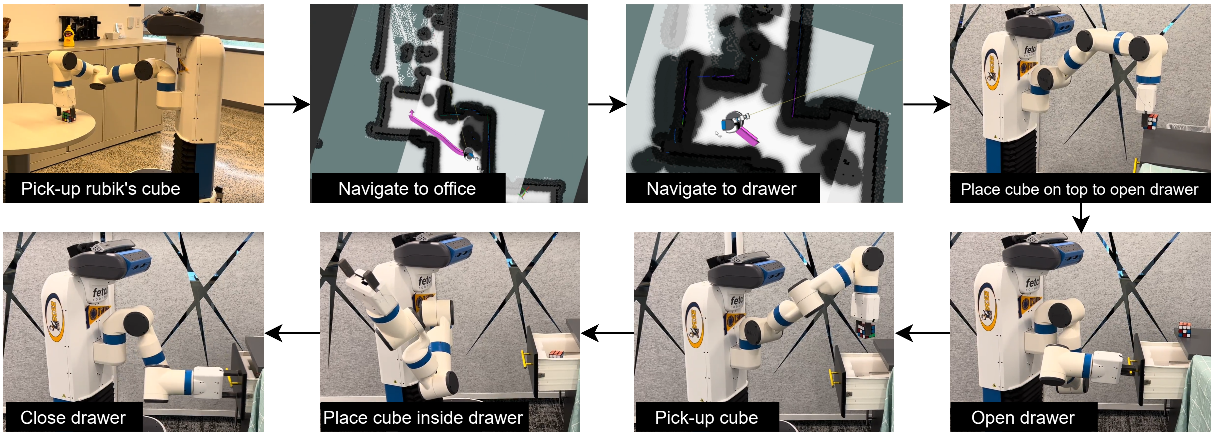

Figure 4.9. Complex Interaction task

4.6 Conclusion

In this work, we explore the intricate development of an autonomous household tidy-up service robot, tackling multiple challenges. Our objective is to create a versatile robot capable of object detection, fine manipulation, and effective navigation in complex real-world environments.

We begin by dissecting the fundamental components of our robot needs for tidy-up task, including semantic mapping, object recognition, and manipulation planning. We emphasize the significance of leveraging both common-sense reasoning and human preference, underpinned by a dataset of human-labeled preferences for object placement. Collaborative filtering methods are introduced, and we integrate Large Language Models to augment our matrix completion approach.

The work further delves into manipulation techniques, highlighting the planning, interaction, and motion aspects. We also touch on navigation capabilities, emphasizing room-to-room and receptacle navigation. The integration of Behavior Trees (BTs) serves as our control architecture, facilitating seamless transitions between tasks. Lastly, we discuss our real-world experiments, including user preferences, long-horizon tasks, and complex interactions, showcasing the robot's adaptability and performance.

This work offers a holistic view of our robot's design, encompassing core components, integration with LLMs, manipulation and navigation strategies, control architecture, and practical experimentation.

4.7 Acknowledgements

I thank Shrutheesh R. Iyer for his work on system integration and control architecture; Anwesan Pal for his work on navigation; Jiaming Hu for his work on robotic manipulation; and Aditya Aggarwal for his work on object and receptacle detection. This chapter contains material from ”Household navigation and manipulation for everyday object rearrangement tasks.” by Shrutheesh R. Iyer, Anwesan Pal, Jiaming Hu, Akanimoh Adeleye, Aditya Aggarwal and Henrik I Christensen, In Submission for the 7th IEEE International Conference on Robotic Computing (IRC), 2023. This dissertation author was one of the primary investigators and authors of this work.

Chapter 5

Conclusion & Future Challenges

This dissertation represents the culmination of several years of dedicated study and research in the field of service robotics, with a particular focus on their applications in organizational tasks. This work draws inspiration from two primary objectives:

Firstly, it addresses the pressing need for robots to assist a diverse user base, including the elderly, those with motor impairments, and the average consumer, in activities of daily living. Enabling greater ease and autonomy in users' lives has far-reaching benefits for society as a whole.

Secondly, for the successful integration of robots into society, it is imperative to systematically test and comprehend the limitations of robotics in real-world scenarios. Conducting research in robotics enables us to pinpoint areas that require exploration, ultimately paving the way for service robotics to evolve into viable products that seamlessly integrate into society.

5.1 Summary of Contributions

5.1.1 Organizing Grocery Items

In the first study presented in this dissertation, we focus on the specific task of organizing groceries in a pantry shelf. This is a routine necessity typically carried out weekly or monthly. With the rise of online grocery delivery, it is a task that could soon be accomplished fully autonomously with service robots.

Our investigation centers on the phase of this task where groceries have been brought indoors and arranged on the kitchen counter, ready to be put away in the pantry. To accomplish this, we employ a methodology that utilizes food ontology data, incorporates common geometric container packaging constraints, and leverages semantic word association as contextual knowledge. This contextual knowledge serves as the foundation for organizing new grocery items within a user's pantry.

Our objective is to maintain the user's existing pantry organization scheme by matching new grocery items with the closest related existing pantry items. We demonstrate the feasibility of this approach by giving our contextual knowledge a mathematical representation and introducing a function that assesses the similarity between two food items. Finally we show through real world experimentation a service robot organizing grocery items in different existing pantry configurations.

5.1.2 Organizing daily objects across a Home

The second study presented in this dissertation focuses on how a service robot can leverage user object location preferences and large language models to identify misplaced objects within a home and reorganize them to their proper locations.

To accomplish this, we adopt a recommender system method known as collaborative filtering and enhance it by incorporating the word embedding space of a large language model (LLM). This approach aims to harnesses the semantic understanding of LLMs to assist our collaborative filtering algorithm in determining users' preferred object placement locations. We showcase a robot system that is capable of navigating from room to room, identifying displaced objects, and returning them to their correct locations.

5.1.3 Joint Organization task for Robots and Humans

In the appendix of this dissertation, we delve into tangential work on service robotics in organizational tasks. Here, we investigate the dynamics of human-robot interaction within tasks that demand dual cooperation. Our primary focus is on the perceived notions of trust and reliability, evaluating how these concepts evolve over time during joint action task and particularly after an initial priming [176].

This supplementary section of the dissertation aims to provide valuable context for both robots and robot designers, not only in the context of joint organizational tasks but also in general collaborative efforts between robots and humans. Our findings have the potential to enhance robot behavior and communication in such tasks, ultimately leading to greater success in overall task completion.

5.1.4 Robots in the Real world

Throughout this dissertation, real-world experimentation stands as a crucial component. Across all the contributions made within this work, I demonstrate the feasibility of real-world robots efficiently executing organizational tasks. I highlight specific capabilities within computer vision, motion planning, and task-level planning that are essential for these tasks. Furthermore, I outline the necessary enhancements required for these skills and their seamless integration. These enhancements encompass enabling robots to reason more flexibly and adapt to goals beyond their predefined plans, bolster their capability to manipulate objects and engage in collaborative tasks and communication with humans, and finally, fostering their ability to understand and interact with novel objects.

These improvements collectively point toward a future where robots can perform organizational tasks, and other functions, in real homes as products for consumers.

5.2 Closing Remarks and Open Questions

This dissertation adopts a bottom-up approach to the challenge of enabling a service robotic system to organize household objects. While our demonstrations effectively showcase how contextual knowledge can significantly improve a robot's efficiency in completing this task, there remain several open areas of research. Specifically, three key areas of interest are centered around robot manipulation, reasoning beyond predefined plans, and interaction with novel objects.

Firstly, it's widely acknowledged within the robotics community that manipulation remains an unsolved problem. While recent developments have seen robots improve their object grasping capabilities, including handling novel objects more successfully than before, we are still distant from achieving human-level dexterity in interacting with objects. This limitation constrains the complexity of organization tasks that robots can perform within homes, including seemingly straightforward tasks like pick and place. Challenges such as the base placement problem for mobile manipulation and non rigid objects, further contribute to this ongoing challenge.

Reasoning beyond predefined plans has been a long sort after goal in robotics and artificial intelligence. Although the work in this dissertation partially abstracts from this problem, in both orginizaial projects, the robot does not have a set goal on where objects should go. It must instead use the knowlege provide to make decisions while running. One limitation in this work however is that once this knowledge is provide, it is not malleable must be updated at before an task begins. Althought the structer of the knowledge proviced allows this to be possible, it was not a focus of this work and therfore was not extended to allow so. This however is an improtant step for robots in the home. Second, while the robot is preforing an action, there is littel room for dynamic changes in the envrioment or goal. Similar to the first limitaiton, while this was not a goal of this work, it is important for robots in the home to be felxible and able to adapt to changes in thier goal or plan in real time.

Achieving reasoning beyond predefined plans has long been an aspiration in robotics and artificial intelligence. While the work in this dissertation does address this challenge to some extent in both organizational studies, there are limitations to our approach. For both projects, the robot does not have a predetermined goal regarding object placement and instead relies on provided knowledge to make decisions in real time. Once knowledge is provided however, it becomes static and cannot be modified during task execution; any updates must occur before the task begins.Although the structure of the provided knowledge theoretically allows for flexibility, this aspect was not the primary focus of this work and was therefore not extended to accommodate such adaptability. Nonetheless, this capability is crucial for robots in domestic settings.

Furthermore, during task execution, the robot's ability to adapt to dynamic changes in the environment or goals is limited. Similar to the first limitation, while this wasn't the primary goal of this work, it is essential for household robots to be flexible and capable of real-time adjustments to their goals or plans in response to changing circumstances.

Lastly, an intriguing open problem in the context of this dissertation pertains to interaction with novel objects. All objects interacted with in this work are assumed to be known and have rigid properties, which is not always the case in real-world home environments. The ability to infer an object's identity from semantic labels, geometric shape information, and general intelligence is a valuable skill for robots as they engage with the world while performing organizational tasks and other task in real homes.

Appendix A

Organizing Objects with Humans in the Loop

Advancements within human-robot interaction generate increasing opportunities for proximate, goal-directed joint action (GDJA). However, robot errors are common and researchers must determine how to mitigate them. In this paper, we examine how expectations for robot functionality affect people's perceptions of robot reliability and trust for a robot that makes errors. Here 35 participants ($n = 35$) performed a collaborative banner-hanging task with an autonomous mobile manipulator (Toyota HSR). Each participant received either a low or high functionality framing for the robot. We then measured how participants perceived the robot's reliability and trust prior to, during, and after interaction. Functionality framing changed how robot errors affected participant experiences of robot behavior. People with low expectations experienced positive changes in reliability and trust after interacting with the robot, while those with high expectations experienced a negative change in reliability and no change in trust. The low expectation group also showed greater trust recovery following the robot's first error compared to the high group. Our findings inform human-robot teaming through: 1) identifying robot presentation factors that can be employed to facilitate trust calibration, and 2) establishing the effects of framing, functionality, and the interactions between them to improve dynamic models of human-robot teaming.

A.1 Introduction

The rapid development of robotics technology is supporting increasing opportunities for humans to work closely with robots across a range of sectors, ranging from housework, to supporting people with disabilities, to aiding people in the workplace, and helping with clinical procedures. [10, 88, 137, 178, 51]. Across these application domains, people are starting to see robots as teammates instead of tools [151, 175, 126].

Human-robot teams are most successful when there is an appropriate level of trust between the human and the robot [116]. In other words, the level of trust a human places in a robot teammate should be calibrated to the robot's abilities. Without such trust calibration, people may over-rely on a robot (e.g., neglecting to effectively monitor its behavior), or under-rely on it (e.g., not use it, or intentionally disable it). Either of these states can degrade the effectiveness of a team during human-robot interaction (HRI) [58]. HRI researchers agree that supporting trust calibration is central to the future of HRI. However, our field currently has a limited understanding of how trust calibration is established and maintained within a human-robot team [58, 34, 54].

As Hancock et al. [58] note, our ability to achieve better trust calibration is dependent on our understanding of the factors that influence the evolution of trust during interaction. In their meta-analysis of existing work on trust in HRI they found that robot reliability is one of the primary factors influencing the development of human trust. It is important to note that human-robot trust-calibration is not only influenced by robot-related factors, but also by a person's expectations. These expectations play a large role in a person's initial experience of a robot [55, 120, 164], critically impacting trust-calibration and future collaboration.

Previous work indicates that informing individuals that a robot has limited capabilities can be advantageous to HRI [120]. However, the interaction in this study was open and not goal-directed and did not explore how a person's expectations might affect their perception of robot capabilities when they are required to engage in collaborative, goal-directed HRI. It seems

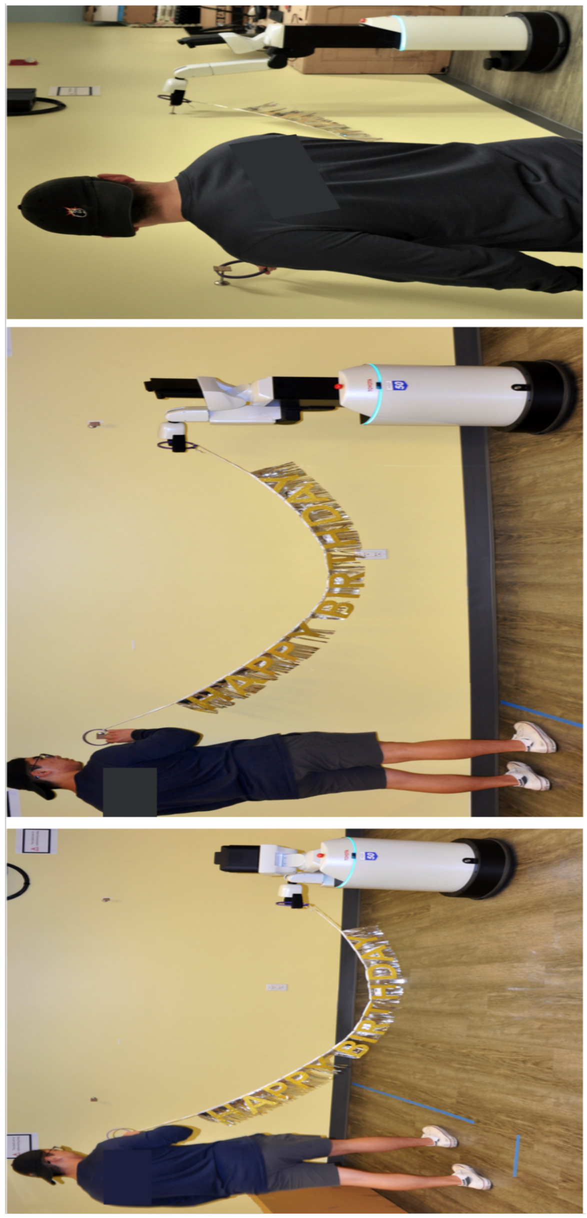

Figure A.1. The three phases of the task. (Left) The robot and participant start at the marked area and then proceed towards the wall. (Middle) The robot looks at the participant to signal it is ready to hang. (Right) The robot then hangs the banner with the participant.

likely that when an individual's expectations for robot behavior are not met the individual will see the robot as unreliable and untrustworthy. Ultimately this may interfere with trust calibration and lead to under-reliance on the robot during human robot teaming.

To our knowledge, there are no studies that examine the influence of functionality framing in the context of proximate, goal-directed HRI. This lack of research is striking given that a wide variety of rapidly expanding HRI domains, including manufacturing, healthcare, home use, and education, require that human-robot teams engage in proximate, goal-directed joint action (GDJA). Not only is proximate GDJA quickly becoming one of the most common modes of HRI, it is also one of the most challenging from a robotics perspective.

At present, robots are prone to many types of errors [68]. The majority of these errors are due the fact that robots have to operate within dynamic environments using noisy sensors [34]. Proximate GDJA contexts present an especially large number of opportunities for robot errors, as humans themselves contribute to the variability of a teaming environment [136, 65, 35, 112, 72, 73, 74].

In this work, we examined the effects of functionality framing on participants' perceptions of robot reliability and trust during proximate GDJA with a robot exhibiting errors. We split participants into two groups and gave them information about the robot with a low functionality framing ("the robot is malfunctioning") or a high functionality framing ("the robot is fully functional"). All participants experienced a robot making errors during a collaborative task (co-hanging a banner, see Fig. A.1), regardless of the functionality framing condition they were in. We formulated four central research questions accordingly:

Does functionality framing affect changes in perceived robot reliability before and after interaction with a faulty robot? Given the previously mentioned association between robot reliability and trust, we examined the effects of functionality framing on perceived reliability following robot errors. We expected that before interacting with the robot, individuals in the low functionality framing group would provide lower reliability ratings than those in the high functionality framing group. We further predicted that the low functionality group would display positive changes in perceived reliability following interaction and the high functionality group would exhibit negative changes.

Does functionality framing affect changes in trust before and after interaction with a faulty robot? We investigated the effects of functionality framing on trust following robot errors in order to improve the current understanding of trust-calibration development in GDJA. We expected that the effects of low vs. high functionality framing on trust ratings for a robot would be similar to those on perceived reliability (i.e., lower pre-interaction ratings for the low vs. functionality group, but also a positive change in ratings for the low group and a negative change for the high group).

How does perceived reliability evolve over the course of interaction with a robot making intermittent errors? We aimed to illustrate how robot errors shape perceptions of robot reliability changed over time to better support the design of robot presentation and behavior for supporting trust calibration. To do this, we assessed the evolution of perceived reliability over the course of interaction for individuals with low vs. high functionality expectations. We predicted that an individual's reports of robot reliability would decrease immediately following robot errors, but that the pattern of recovery following each error might vary depending on the functionality framing they had received.

How does trust evolve over the course of interaction with a robot making intermittent errors? We evaluated the evolution of a person's trust over the course of interaction with the robot in our GDJA task. As for perceived reliability, we expected that robot errors would be followed by decreases in trust, but that the pattern of recovery following each error might be influenced by an individual's functionality framing condition.

Our four research questions were designed to generate an understanding of the relationships between functionality framing, robot errors, perceived reliability, and trust that would allow us to make recommendations for the use of functionality framing to support trust calibration during human-robot teaming. We expected that our functionality framing would have similar effects on both perceived reliability and trust. However, we also predicted that there would be some differences between participants' reports of perceived reliability and trust. To evaluate the convergence and divergence of perceived reliability and trust during our GDJA task we compared the findings generated by the first and second pairs of questions (i.e, Question 1 and Question 2; Question 3 and Question 4). The results of the first pair of questions allowed us to asses the similarity between changes in trust and perceived reliability for pre- vs. post-interaction questionnaire responses while the outcomes of the second pair of questions provided information about similarities and differences in the evolution of perceived reliability and trust during ongoing interaction.

Our results indicate that functionality framing about robot performance does significantly impact both perceived reliability and trust for a robot co-actor. Individuals with low expectations for robot functionality exhibited a positive change in both reliability and trust questionnaire responses following interaction.

The contributions of this work are as follows. First, we show that functionality framing impacts how people's expectations in robot behavior align with their experience during interaction through effects on both perceived reliability and trust. Second, we illustrate how functionality framing shapes the actual evolution of human-robot teaming through effects on perceived reliability and trust. This work therefore presents opportunities for transformative advances for proximate HRI by: 1) identifying robot presentation factors that can ultimately be employed to facilitate trust calibration, and 2) establishing the effects of framing, functionality, and the interactions between them in order to improve dynamic models of human-robot teaming.

A.2 Background

A.2.1 Robot Expectations

Humans bring many attitudes, beliefs and previous experiences with technology to their interactions with unfamiliar robots. Together, these factors act to shape the expectations an individual has for a robot prior to interaction. Expectancy Confirmation Theory suggests that expectations consist of the characteristics a person anticipates that a product, service or technology artifact will have [117, 118]. Expectations are said to directly influence both perceptions of performance and disconfirmation of beliefs for an entity.

As a result, a person's satisfaction with a product, service, or technology artifact following interaction is directly influenced by their perception of its performance and disconfirmation of their prior beliefs. In summary, there is an relationship between pre-adoption expectations and satisfaction that is mediated by the disconfirmation of beliefs.

Robot design also influences people's expectations and consequently affects how they evaluate the robot after interacting with it [36, 47, 127]. This is especially true with respect to the humanness of a robot's appearance.

When the social expectations elicited by humanoid robots are met, people experience interactions with them as enjoyable and empowering [131]. Conversely, users are sometimes disappointed and dissatisfied when their expectations for human capabilities are not met [38, 108]. This kind of miscalibration between human expectations and robot capabilities based on robot appearance has recently been referred to as the Form Function Attribution Bias (FFAB) [59].

A.2.2 Framing

User experiences are also shaped by any information they receive prior to HRI about the kind of behaviors a robot is intended to perform [55, 164]. The presentation of such information is either referred to as framing (e.g. [55, 129]), or expectation setting (e.g. [120, 164]). Researchers in psychology suggest that framing or expectation setting helps people to process new information through the activation of contextually relevant knowledge [169, 168]. For example, prior work suggests that giving drones a prevention frame as opposed to a promotion frame leads to increased public support [129]. In that study, people were more supportive of drones if they are presented as "capable of protecting people from harm" vs. "actively seeking out illegal activities to support increased prosecution and punishment".

Groom et al. [55] suggest that framing in proximate HRI is often as simple as sharing a few facts about a robot prior to interaction. In their work, they found that this process led to greater compliance and confidence in robotic agents during a simulated search and rescue scenario. The authors interpreted this increase in ratings as signifying an increase in trust for framed robots.

Generally positive portrayals of robot abilities have been shown to result in user experiences of dissatisfaction and disappointment for task-oriented robots [120]. In this context, researchers used functionality framing to present a robot as either limited or skilled in a set of abilities. Individuals who expected the robot to be limited exhibited an increase in their perception of robot capabilities following interaction. In contrast, individuals who expected the robot to be skilled displayed a decrease in their perception of robot capabilities after interaction. This outcome of dissatisfaction is predicted by Expectancy Confirmation Theory across contexts where a person's initial expectations are too high compared to their actual experience [117, 118].

Prior work shows that the expectation for automated aids to perform at nearly perfect rates leads operators to pay too much attention to errors and ultimately underestimate the reliability of the system [40]. This can lead to unwarranted distrust and disuse. Instead, underrepresenting a robot's capabilities either through its morphology or the information included in its presentation may provide the greatest potential for successful human-robot teams. In the current study, we evaluate the effect of using framing to transparently or falsely communicate the current functionality of a robot.

A.2.3 Reliability and Trust

Through a meta-analysis of factors affecting trust in HRI, Hancock et al. [58] demonstrated that performance-based robot characteristics have the largest influence on robot trust. As noted in this meta-analysis, trends in the literature suggest that higher levels of trust are associated with greater robot reliability. Previous work suggests that human trust during HRI is dynamic [142, 123]. This has been especially apparent in the context of interaction with a robot that makes errors [146, 173, 140].

Figure A.2. In our study, a robot and person engaged in proximate, GDJA to collaboratively hang a banner. Intermittently, the robot intentionally dropped the banner 7 cm short of the hook. The robot would still proceed to the "target" location as if it were still trying to hang the banner.

A.2.4 Robot Errors

Robot errors are common when robots operate in unstructured environments, which include objects and individuals that may change frequently change location or behave in other ways that are hard to predict [57, 140, 173, 68]. While it is important to continue making robots more reliable, the complexity of human behavior makes eliminating errors during proximate HRI virtually impossible. The more we know about how robot reliability affects human-robot teaming, the better we will be at achieving effective proximate HRI even when a robot makes errors.

Rovira et al. [143] showed that human-robot team performance on a navigation decision task was already superior to human performance alone even when the robot operated at 80% reliability. This is consistent with previous work showing that the performance of a human-automation team will be more efficient than a human completing the same task without automation as long as the reliability of the automation is greater than 70% [179]. In cases where 100% reliability cannot always be guaranteed, it is important to explore the effect that robot errors have on user perceptions.

Previous work showed that robot errors significantly impact a human teammate's attributions of robot trust and processes of human-robot trust calibration across HRI [159]. Thus, robot reliability plays a substantial role in shaping people's trust during HRI [58]. Existing work suggests that continuous robot errors can lead to decreases in human trust. However, if robot reliability is high for a period of time at the beginning of an interaction then a robot error may have a less severe effect on trust [48, 91].

Through the use of a real-time trust measure, Desai et al. [33] similarly observed that periods of low robot reliability early on in an interaction had more of an impact on trust than periods of low reliability occurring closer to the middle or end of an interaction. The measurement of trust evolution over the course of an instance of HRI therefore provides valuable information about the impact of robot performance on human experience. Understanding this evolution also allows us to predict future events such as the transfer of trust between HRI tasks [159].

The magnitude of trust decrease following an error, as well as the duration of subsequent trust-recovery, are impacted by the severity of the error along with the frequency of past errors [29]. An individual's perception of error severity is, in turn, likely determined by the task and context. For example, in a goal-oriented task robot errors are especially undesirable and cause users to perceive the robot as less reliable and trustworthy [146]. In time-critical situations, like emergency evacuations, most users completely distrust a robot after a single error [141].

A number of previous studies also indicate that people find robot errors less egregious if the robot acknowledges the error in a way that it is consistent with a typical human response, or is human-like in its appearance or speech [23, 140, 146, 92, 60]. However, because robots do not always know when they have made an error, such acknowledgment is not always possible. There are also situations in which the robot is aware of an error but still cannot explain why the error occurred. Whenever it is possible that a robot will not be able to acknowledge its own errors it is especially critical to provide appropriate framing regarding robot functionality prior to HRI.

A.2.5 Current Study

In the current study, we compared the effects of pre-interaction framing for low vs. high robot functionality. To do this we evaluated participants' perceived reliability and trust using measures which take into account multiple timescales of experience. As a result, we are able to establish the effect of robot errors on constructs central to human-robot teaming within the context of a GDJA task.

A.3 Methodology

In this study, we provided participants with information indicating that the robot's arm was either fully functional or malfunctioning. Then we asked them to hang a banner with the robot (see Fig. A.1). We programmed the robot to make the same errors each time the experiment was performed, regardless of the participant's framing condition. This experimental design allowed us to establish the effects of an individual's expectations about robot functionality on trust and perceived reliability in the context of robot errors.

A.3.1 Study Design

For our experiment we employed a Toyota Human Support Robot (HSR). This robot has a mobile base and an arm with five degrees-of-freedom. It stands 1.35 meters tall and has 12 different sensors, including a stereo camera, two wide-angle cameras, a RGB-D sensor, joint angle encoders, a potentiometer, two force sensors, an IMU, a magnetic sensor, and a LIDAR. The HSR was fully autonomous throughout the current study.

Each participant performed 10 experimental trial trials. In a single trial the participant and robot moved toward the wall and hung a 7-foot long banner together. The robot was programmed to hang the banner using predetermined locations and joint configurations. We produced a map of the environment using Hector Mapping [85]. At the beginning of an experimental session, we set the robot's initial starting location via RViz.

Before the start of each trial, a member of the research team placed the banner in the robot's gripper. Once the trial began, the robot moved forward until it reached a preset distance from the wall (Target Location 1). It then lifted its torso, rotated its head toward the participant, and raised its arm. After a brief pause, the base continued to move to a second preset location (Target Location 2). When it reached this location the robot opened its gripper to release the banner onto the hook.

In trials three and seven, the robot made a pre-programmed error: it dropped the banner on the floor (see Fig. A.2). To perform the error, the robot exhibited the same behavior as in successful trials until making a premature stop 7 cm short of Target Location 2. It dropped the banner at this position, and then proceeded to Target Location 2 to give the impression of an unintentional error.

We included two error trials, out of the 10 total experimental trials per participant. Thus, the robot's reliability reflected the 80% reliability found to support effective collaborative performance in previous human-automation teams [143]. We introduced the first error in trial three in order to allow participants to build an initial baseline level of trust before the robot made any errors [33]. We placed the second error in trial seven so that we had an equal number of recovery trials following each error (i.e., three) in which to assess the evolution of perceived reliability and trust.

The robot was able to talk through text-to-speech synthesis. It said "starting now" to signify the start of each trial, and "hanging complete" to signify that it had released the banner for hanging. The robot's speech was consistent across all experimental trials. On successful trials the robot made the "hanging complete" announcement after it had hung its end of the banner the wall. In the error trials the robot made the announcement after it had dropped the banner on the floor.

The robot did not exhibit any error acknowledgment. As noted earlier, prior work suggests that robot acknowledgment can mitigates the effect of errors [68, 92]; however, in our study, any error mitigation resulting from robot acknowledgment might have influenced participants' perceptions of robot errors. This would have interfered with our ability to evaluate the effect of functionality framing on the perception of errors. In excluding the possibility of error mitigation in the current study, we are able to provide an initial baseline understanding of the effects of functionality framing on people's experience of robot errors in goal-directed HRI.

A.3.2 Variables

Prior to the experiment, participants received information about the robot including details intended to bias their expectations about robot functionality. Participants in the "low" functionality framing condition received a sheet stating that "The robot's arm is malfunctioning today," accompanied by a warning symbol. Participants in the "high" functionality framing condition received a sheet stating "The robot's arm is fully functional today," accompanied by a check symbol. We also posted signs on the target wall for the banner hanging task reiterating the framing information corresponding to each participant's condition, consistent with prior HRI work on framing outside of GDJA contexts [120, 164]. We randomly assigned participants' functionality framing condition within groups of 10 participants such that each group included five individuals in the low condition and five in the high condition.

Prior to running the current study, we conducted a pilot study on Amazon Mechanical Turk to assess the effects of functionality framing information on perceived reliability and trust. We asked participants to make evaluations based on the hypothetical context of completing a collaborative banner-hanging task with the HSR (see Section A.3.3 for details on the questionnaire instruments employed). We used this pilot to establish the effectiveness of the framing language and presentation used in the current study.

A.3.3 Measurements

We used validated questionnaires to assess perceived robot reliability [83, 156] and trust in the robot [109] before and after interaction with the HSR. We also obtained real-time measures of reliability and trust following each experimental trial.

Each validated questionnaire consisted of four items (see Table A.1). We asked participants to respond to these questionnaires before they interacted with the robot (i.e., directly after the manipulation check) and again after they completed all of the interactive experimental trials. Before interaction with the robot, we prefaced each questionnaire with the statement "Imagine you're going to hang a banner with this robot. Based on the information you just learned, please respond to the following questions." After participants completed the banner-hanging trials, we introduced each questionnaire with the language "Now that you have finished performing the task with the robot, please update your ratings.".

We used a reliability questionnaire consisting of four items [83, 156]. We provided response options to these items as discrete visual-analogue scales (DVASs) with six options. The scales ranged from "Strongly Disagree" to "Strongly Agree", with only the endpoints labelled. We also used a trust questionnaire made up of four items [109]. We also provided DVASs for these items with response options ranging from "Not at all" to "Completely" with only the endpoints labeled. We evaluated each questionnaire by averaging the numeric values corresponding to an individual's response selections for each of the four items. High scores indicated greater reliability and trust on the respective questionnaire instruments.

Using the real-time measures of reliability and trust administered following each experimental trial, we generated an event contingent series for the evolution of each construct over the course of interaction. These series are similar to those obtained through Ecological Momentary Assessment (EMA) paradigms [153]. For these real time measurements we had participants indicate whether their reliability and trust in the robot were increasing ($\uparrow$), decreasing ($\downarrow$), or remaining the same ($\leftrightarrow$) after each iteration of the banner-hanging task. We adapted this response measure from one designed by Desai et al. [33]. We administered it here via paper with reliability presented on the top half of a single sheet of paper, and trust presented on the bottom half. There was also a place for participants to write comments associated with each rating.

It is important to note that we were interested in generating information about people's perceived reliability and trust for the robot based on their own internal conceptions of each construct. In order to do this we did not define “reliability” and “trust” for participants at any point during the study.

After participants completed the post-interaction reliability and trust questionnaires, we asked them to complete a questionnaire on their willingness to collaborate with the robot on a number of additional tasks (see Conclusion). We created this questionnaire based on tasks Beer et al. [10] had previously identified as tasks for which older adults would prefer robot assistance to human assistance. This consisted of 14 home tasks, 10 of which are primarily physical (cleaning a kitchen, cleaning a bathroom, cleaning a window, pest control, loading dishes in a dishwasher, making the bed, cleaning the floor, taking out the trash, doing laundry) and four of which are solely cognitive (reminders to take medication, staying informed on weather/news, learning how to use technology, getting information about your hobbies). For each task, participants indicated their willingness to collaborate with the robot using a six-option DVAS, ranging from “Not at all” to “Completely”, with only the endpoints labeled. While we recognize that this measure has not undergone a reliability analysis, we thought that it could still provide valuable information in the context of the current study.

A.3.4 Procedure

When a participant arrived the experimenter first brought them to a conference room close to the lab. The experimenter then provided the participant with information about the study and the robot. First, the participant received a single page entitled “Human Support Robot (HSR)”, which had a picture of the robot and the information that it was an autonomous robot designed to support human activities. On this page we provided three bullet points outlining the robot’s ability to : 1) “Sense objects and people around it”, 2) “Grab/pick up objects using its gripper arm”, and 3) “Move around a room”. We confirmed that the participant did not have any questions about this information and then presented a second page, which included the functionality framing information about the robot consistent with the individual’s randomly assigned condition. As we describe in Section A.3.2, we told participants in the low functionality

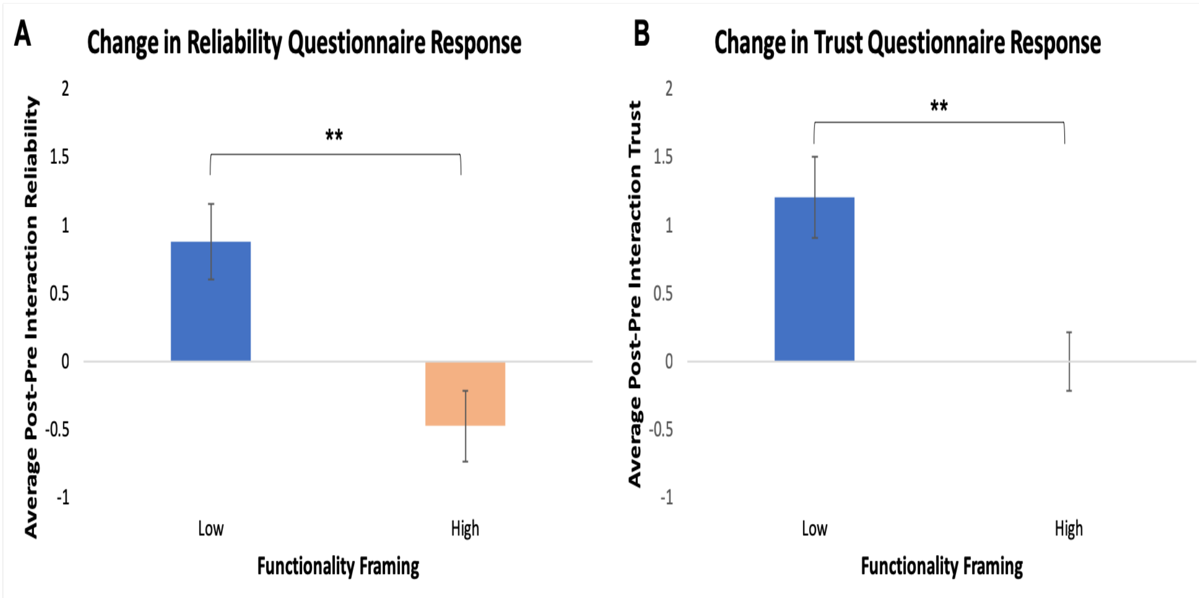

Figure A.3. Average change in reliability (A) and trust (B) for post vs. pre-interaction in the low and high functionality framing groups. Error bars show standard error. $^{**}p < .01$.

condition that the robot's arm was malfunctioning, and those in the high functionality condition that the robot's arm was fully functional.

After the participant completed a manipulation check (see Section A.4.1 for details), the experimenter gave them copies of the reliability [83, 156] and trust [109] questionnaires used in the current study. To avoid any influence of the experimenter's presence on participant responses, the experimenter asked the participant to complete the questionnaires alone and meet them in the lab when they were finished.

When the participant entered the lab, the experimenter explained the collaborative GDJA task in view of the robot and experimental test-bed. The banner-hanging task required that the participant and robot each hold one end of the banner, move together across the room to the designated target wall, and hang their respective ends of the banner on existing hooks. The experimenter told the participant that before each trial began, another member of the research team would hold one end of the banner close to the robot's gripper to initiate robot grasping. The researcher controlled this initial grasping action with a two-key key press via a Bluetooth keyboard. The researcher then handed the other end of the banner to the participant. After leaving the test-bed and confirming that the area around the robot was clear of obstacles, the researcher allowed the robot to begin the trial via another two-key press. (After which the robot was fully autonomous). The participant had an opportunity to do one preliminary practice trial and ask the experimenter any questions they had about the task.

The experimenter then requested that the participant answer the real-time response items addressing perceived reliability and trust after each of the remaining experimental trials. Once the experimenter confirmed that the participant did not have any remaining questions about the task, they explained that the other member of the research team would oversee the preparation of the robot for each trial. The original experimenter then left the room for the duration of the experimental trials. We made this experimental design choice in an effort to 1) reduce the participants' experience of being observed and evaluated during the task and 2) minimize any influence of human-human interaction outcomes on participants' perceptions of robot behavior.

Once the researcher initiated a trial the robot would say "starting now" before starting to move autonomously toward the target wall. Participants started each trial facing the wall then walked towards it with the robot to hang the banner. Upon reaching the wall, the robot would lift its torso and arm then look toward the participant. We asked the participant to wait until the robot looked toward them to hang their end of the banner. Once the robot completed its hanging movement sequence for the trial it would say "hanging complete" before turning around and returning to the starting position.

While the robot returned to the starting position the participant made their real-time reports about whether their perceived reliability and trust for the robot were increasing, decreasing, or remaining the same. Participants made these reports via a sheet of paper on the edge of the test-bed. The robot made the "hanging complete“ statement in all trials, including trials three and seven in which it made a pre-programmed dropping error (see Section A.3.1 for details). When the robot made the first dropping error, the researcher asked the participants to leave the banner on the floor for them to retrieve.

Once a participant had completed the 10 experimental task trials, the researcher gave them copies of the four-item reliability and trust questionnaires for the second time. The researcher then asked the participant to complete the *Willingness to Collaborate* questionnaire. Lastly, the researcher collected demographic information and the researcher and experimenter debriefed participants about the true nature of the study. A single study session took around 30 to 40 minutes to complete.

In total, 35 individuals participated in the current study. All were fluent English speakers. Seventeen individuals experienced the low functionality framing condition and 18 experienced the high functionality framing condition. These sample sizes were chosen as consistent with the majority of existing work in HRI. Participants ranged from 18 to 29 years of age ($M = 26.66$, $SD = 3.00$). Twenty-four participants identified as women and 11 identified as men. Twenty-six participants reported being at least "somewhat familiar" with robotics. All participants provided informed consent. All of the procedures and data collection tools used in the current study were approved by our institutional IRB.

A.4 Results

A.4.1 Manipulation check

Each participant in the current study was asked to complete a manipulation check regarding their functionality framing condition before they entered the lab. After we presented the introductory information about the robot we asked each participant to respond to the questions "Is the robot fully functional today?" and "If no, what's wrong with it?". All participants correctly answered the manipulation check questions based on their framing condition.

A.4.2 Pre- vs. post-interaction reliability and trust

We used [156]'s four-item reliability questionnaire and [109]'s four-item trust scale to assess participants' perceived reliability and trust both before and after completing all 10 experimental banner-hanging trials with the robot. Based on a participant's response to each of

Table A.1. Trust and Reliability Questionnaires

Trust Questionnaire
1. I can depend on the robot to work correctly every time.
2. The robot seems reliable.
3. If I did the same task with the robot again it would do it the same way.
4. I could trust the robot to work whenever I might need it.
Reliability Questionnaire
1. To what extent can the robot's behavior be predicted from moment to moment?
2. To what extent can you count on the robot to do its job?
3. What degree of faith do you have that the robot will be able to cope with similar situations in the future?
4. Overall how much do you trust the robot?

these measures we calculated their average pre- and post-interaction reliability and trust scores. We then used these pre- and post-interaction scores to calculate the change in an individual's reliability and trust following interaction by subtracting their pre-interaction scores from their post-interaction scores. We averaged these change scores by functionality framing group so that we could compare the effect of functionality framing on changes in perceived reliability and trust following interaction.

A paired t-test revealed a significant effect of functionality framing on reliability change scores, $t(33) = 3.58$, $p = .001$, $Cohen'sd = 1.25$. Here, participants in the low functionality condition reported an increase in the reliability of the robot following interaction while those in the high functionality condition reported a decrease (see Fig. A.3A).

A paired t-test on trust change scores also revealed a significant effect of functionality framing, $t(33) = 3.32$, $p = .002$, $Cohen'sd = 1.15$. Participants in the low functionality condition reported an increase in trust following interaction while participants in the high functionality condition reported no change in trust (see Fig. A.3B).

A.4.3 Real-time reliability and trust

We also asked participants to provided real-time, event contingent reports of reliability and trust during interaction with the robot. Participants completed these assessments after each

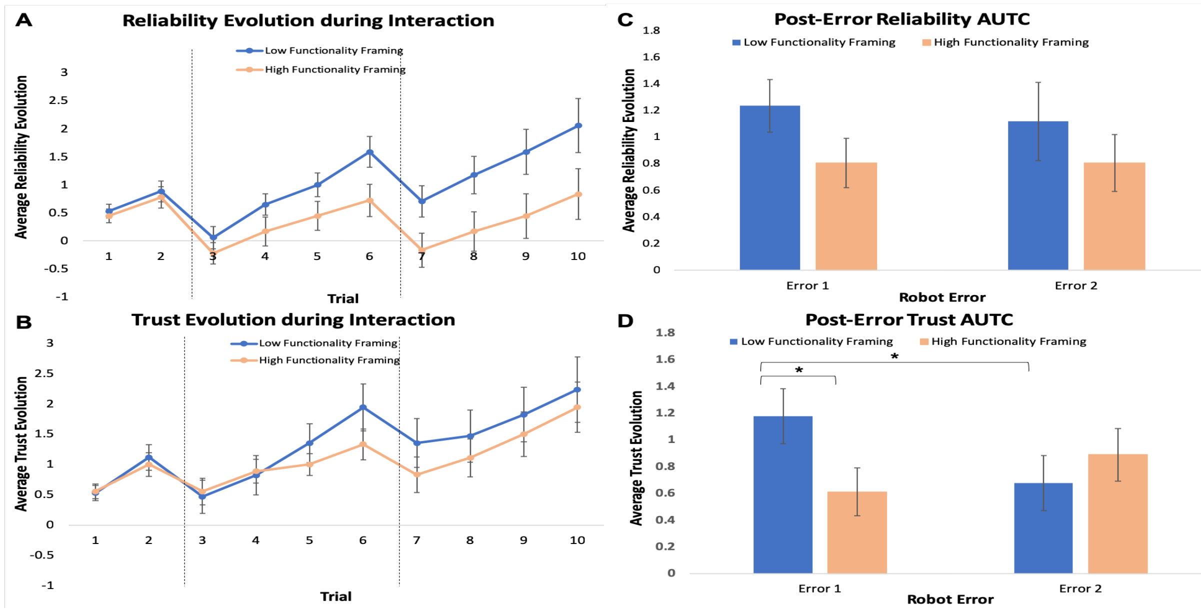

Figure A.4. Evolution of reliability (A) and trust (C) during interaction for the low and high expectation groups. Dashed lines show trials in which the robot made a banner-dropping error (Trials 3 and 7). Average reliability (B) and trust (D) evolution following each of the two robot errors for participants in the low and high funcEtionality framing conditions. Error bars show standard error.

experimental trial, indicating whether the reliability of the robot and their trust in the robot were increasing, decreasing, or remaining the same [33]. We initially assigned directional values to each response category, using +1 for "increasing", -1 for "decreasing", and 0 for "remaining the same". We then used these directional values to estimate separate construct evolution series for reliability and trust as they changed over the course of an individual's interaction with the robot. The first value of each construct evolution series was the directional value corresponding to a participant's response for that construct (i.e., perceived reliability or trust) after the first experimental trial. We then calculated all subsequent values in the series by finding the sum of the directional value for the current trial and the construct evolution value for the previous trial (e.g., if the participant's trust was "remaining the same" after the first trial and "increasing" after the second trial, then the first two values of the construct evolution series for trust would be 0, and 1).

The evolution of real-time perceived reliability and trust can be seen in Figs. A.4A,C (respectively). We statistically evaluated the effect of framing condition on each of perceived reliability and trust evolution by treating each error trial as an anchoring event [153] and evaluating changes in the three trials following that error. To do this we calculated the normalized area under the curve (AUTC) [33] for each construct over each time period of interest (i.e., trials 4-6 for Error 1 and trials 8-10 for Error 2) using the trapezoidal rule (see [62]. We then used an Analysis of Variance (ANOVA) to compare the AUTC for each framing group at each of the timepoints, as is common practice within the analysis of EMA data [16].

A 2 (functionality framing: low, high) $\times$ 2 (robot error: 1, 2) mixed model ANOVA on the AUTC for perceived reliability evaluation revealed no significant interaction between the two variables or main effects (see Fig. A.4B). However, a 2 (functionality framing: low, high) $\times$ 2 (robot error: 1, 2) mixed model ANOVA on the AUTC for trust evolution did reveal a significant interaction between framing and robot error, $F(1,33) = 5.70$, $p = .02$, $\eta p^2 = .15$ (see Fig. A.4D). Pairwise comparisons confirmed that following Error 1 trust increased significantly more in the low vs. high functionality group ($p = .045$). Pairwise comparisons also showed that within the low functionality group there was a significantly greater increase in trust following the Error 1 compared to Error 2 ($p = .01$). These results demonstrate that the interaction between functionality framing and robot error number was driven by the fact that participants in the low functionality group showed a greater increase in trust compared to the high functionality group after Error 1, but not after Error 2.

This interaction was driven by the fact that reliability increased more for participants in the low expectation setting group than the high expectation setting group following the first robot error, but not the second robot error.

A.4.4 Collaboration transfer

After participants completed all ten trials of the banner-hanging task and the post-interaction reliability and trust questionnaires,they responded to our 14-item *Willingness to Collaborate* questionnaire (see Section A.3.3 for details). For each participant, we generated a summed measure for each of the two task domains represented in the questionnaire (i.e., physical vs. cognitive), with a maximum possible sum of 60 for physical tasks and 24 for cognitive tasks. We then divided each participant's sum in each task domain by these possible maxima to obtain their percent willingness to interact in the future. We averaged this percent willingness across participants in each framing condition. This allowed us to conduct paired t-tests comparing the framing conditions within each domain.

There was no significant effect of framing group on willingness to collaborate on either physical ($t(33) = .22$, $p = .83$) or cognitive tasks ($t(33) = 1.5$, $p = .14$). The average willingness to collaborate on additional physical tasks was 73.04% for the low functionality group (SD = 3.91%), and 74.19% for the high functionality group (SD = 3.58%). The average willingness to collaborate on additional cognitive tasks was 70.97% for the low functionality group (SD = 5.96%), and 82.18% for the high functionality group (SD = 4.51%).

A.5 Discussion

HRI researchers, industry professionals, and consumers are all eager to have robots that can engage in proximate HRI within human environments. Unfortunately, human behavior is largely unpredictable, and these environments change frequently and rapidly. These conditions are challenging for robots, leading to problems ranging from sensor occlusion to unsuccessful behavioral planning [68]. As a result, robot errors are common in proximate human-robot joint action. While it is important to continue improving robot resiliency to the challenges of proximate HRI, it is also critical that we learn how to support effective human-robot teaming in the context of robot errors.

In our study, we experimentally manipulated individuals' expectations for the functionality of a robot. We then addressed four specific research questions about the effects of a person's expectations on their perceptions of a robot's reliability and trustworthiness after the robot made errors. Our first research question was whether functionality framing would affect changes in perceived reliability following collaborative interaction with the robot. Our findings revealed that individuals with low expectations thought the robot was more reliable after they interacted with it than they did before they interacted. In contrast, those with high expectations thought the robot was less reliable after they interacted with it.

Our second research question was whether functionality framing would affect changes in trust following collaboration with a robot making errors. Here we also found that the change in participants' post- vs. pre-interaction questionnaire responses was affected by functionality framing. Individuals in the low functionality framing group showed an increase in trust ratings following interaction with the robot, while individuals in the high functionality framing group showed no change in trust.

The collective findings from our first two research questions showed that individuals in the low functionality group displayed positive changes in both perceived reliability and trust following interaction while individuals in the high functionality group showed a negative change in reliability and no change in trust. These results are consistent with previous empirical work showing that human-robot trust is strongly associated with robot reliability [58]. However, our findings also support the idea that trust is robust to some amount of change in robot reliability over time [33]. We expect that the relationship between robot reliability and trust during HRI is largely shaped by the nature of the task. For example, the participants in our study reported that the robot's reliability rate was acceptable for the banner-hanging task, but would not have been acceptable for a more delicate task. They communicated this with statements like the following, which came from three unique participants:

"The robot is overall reliable, since it only failed twice. However, I wouldn't trust it with delicate objects, such as glass, unless it makes zero errors."

"The robot could be used for simple/non-delicate tasks just fine with reliability as it is."

"It would be harder to trust the robot if there was a delicate object involved."

In response to research questions one and two, we predicted that individuals with low expectations for robot functionality would experience more positive changes in their perceptions of the robot following interaction compared to individuals with high functionality expectations. This prediction was confirmed by individuals' pre vs. post-interaction questionnaire responses for perceived reliability and trust. We suggest that anyone in the position of presenting a robot for proximate, goal-directed HRI to new users should be careful not to overstate the robot's capabilities, as this can lead to negative changes in perceived reliability and trust during teaming. Relatedly, since robot humanness appears to induce high expectations for functionality [38, 108] we recommend that designers and roboticists consider robot morphology very carefully. This recommendation supports our prior work suggesting that that the use of humanoid portrayals shapes human expectations in a way that could constitute manipulation, and that the associated ethical implications should be taken seriously [135, 138].

In our third and fourth research questions we asked whether functionality framing would influence the evolution of perceived reliability and trust during interaction. To address these questions, we asked participants to make real-time reports about their perception of the robot's reliability and their trust in the robot after each iteration of the banner-hanging task. These reports revealed that participants' perceived reliability and trust for the robot increased across trials in which the robot performed the task successfully, and decreased immediately following each of the two pre-programmed errors in robot behavior. This trajectory was similar across framing groups.

When we evaluated our third research question, we saw no effect of framing group on the evolution of individuals' perceived robot reliability. This suggests that functionality framing may not affect an individual's perception of robot reliability on a trial-by-trial basis. However, when we evaluated our fourth research question we observed an interaction between functionality framing and trust. Individuals in the low functionality framing group reported a greater increase in trust following the first robot error as compared to the second robot error. In contrast, individuals in the high functionality framing group reported greater trust increases after the second error compared to the first error.

Notably, these findings indicate that individuals with low expectations exhibit greater trust recovery following an initial robot error compared to those with high expectations. Previous work in robot tele-operation has revealed that a user's trust typically increases faster during an initial stable period of high robot reliability than it does after a drop in robot reliability [33]. Our own results suggest that low expectations may mitigate the negative effect that a single robot error has on trust building. Despite our observation that functionality framing has significant effects on perceived reliability and trust before, during and immediately following interaction, functionality framing did not appear to have an effect on the willingness of individuals to collaborate with the same robot on future tasks.

Interestingly, several participants across both functionality framing conditions reported that after the robot made an error they began to closely analyze its behavior. Participants wanted to understand why an error had occurred. They also sometimes thought they could predict when errors were going to occur. While some participants were accurate in identifying factors associated with robot error, many were not. Specifically, some individuals noticed that if the robot raised its arm and then moved forward and stopped short of the normal location where it released the banner (i.e., Target Location 2) it was likely to immediately drop the banner on the floor. Ultimately, the change in robot behavior associated with the error could not be detected until just a few seconds before the robot dropped the banner.

Still, some individuals attributed robot errors to their own behavior early on in a trial (e.g., walking too fast with the banner before reach the robot's first stopping location). This feeling of responsibility for robot errors may be due to the fact that the human and robot had very similar roles in the banner-hanging task. Past work suggests that people feel more responsible for a task when working with a peer or subordinate robot than when working with a supervisor robot [66]. As a result, people may be more likely to assume they have an effect on robot errors in tasks where human and robot roles are very similar compared to those in which the robot has a more subordinate role. Ultimately, the participant feedback we received in the current study emphasized that providing information about the cause of robot errors during HRI may reduce the attentional resources a human co-actor uses in response to robot errors.

A.5.1 Limitations and future directions

Our participants did not experience any acute risk during interaction with the robot. This may limit our ability to contribute to a common discussion around risk and trust in HRI. Many models of trust within HRI include risk as an important component in trust development. These models are often based on a definition of trust offered by Mayer and colleagues [101]. In this definition, trust is based on a person's willingness to be vulnerable to the actions of another party irrespective of the ability to control that party's actions.

The current study requires that individuals engage in human-robot teaming in order to complete a GDJA HRI task. Human-robot teaming itself requires a human to depend on a robot to complete a portion of the collective task. More specifically, the human cannot control the robot's actions and the team's success is vulnerable to the robot's behavior [95]. As a result, our participants experienced a low level of vulnerability during interaction, but no explicit risk. Future studies of human-robot trust during GDJA will benefit from the thoughtful introduction of additional risk to participants.

In order to identify how functionality framing information might impact human-robot team performance in the future we assessed participants' willingness to collaborate with the robot on additional tasks following the completion of the banner-hanging experiment. It is possible that our ability to capture the effect of functionality framing on future collaboration was limited by our use of a self-report measure. Previous work suggests that self-report statements do not always fully represent how an individual would actually act in a given situation [24]. Alternatively, implicit, behavioral metrics can allow researchers to measure constructs, such as trust, frequently and unobtrusively during ongoing interaction (e.g., see Hayes et al. [60]). In future work, researchers may be able identify a relationship between functionality framing and willingness to collaborate using an implicit measure of willingness to collaborate.

It is also possible that we observed similar responses between framing groups on our willingness to collaborate questionnaire because interaction experience is more important than functionality framing in determining a person's willingness to collaborate on future tasks. In other words, participants may have made decisions about collaborating with the robot in the future based on what they learned about the robot's capabilities during interaction instead of the framing information they received about the robot at the beginning of the experiment.

Interestingly, our participants spent considerable time thinking about potential causes for the robot's errors. We chose not to include error acknowledgment in this study as this might have mitigated the effect of robot errors on participants' perceptions of reliability and trust. This allowed us to provide a baseline illustration of the relationship between functionality framing and error during proximate HRI. However, previous work suggests that robot transparency surrounding errors is beneficial to human experience and team performance during HRI [33, 145, 172, 67] and feedback from our own participants indicates that this would be valuable within proximate, GDJA as well. Thus, future work on human-robot teaming fluency in proximate, GDJA should aim to establish how robot responses can reduce the attentional resources humans give to robot errors when they do occur.

A.5.2 Conclusion

Our results suggest that in HRI contexts where robot errors are likely, such as proximate GDJA, human co-actors are likely to be disappointed by robot capabilities unless they are told ahead of time that the robot may make errors. We also showed that when people expect robot errors they exhibit greater trust recovery after a single robot error. Our measures revealed a strong relationship between trust and reliability, while also indicating that there is independence between the two constructs.

Participants in the current study tended to generate explanations about why the robot made errors and to predict when the next error would occur. This observation supports the idea that future research should continue to examine how robots can acknowledge errors during interaction in order to minimize human attention to them. As we have noted, future work should also systematically explore the effect of risk on reliability and trust in proximate GDJA. In conjunction with our current work on robot presentation, these avenues will greatly expand our ability to support trust-calibration across a range of proximate HRI contexts in which robot errors are known to be frequent and impactful events. At present, our findings can be utilized for the continued improvement of 1) framing factors for robots to be used in proximate HRI, and 2) modeling of human-robot trust for the support of effective human-robot teaming.

Willingness to Collaborate Questionnaire

Physical Tasks
1. Do you trust the robot to help you change a light bulb?
2. Do you trust the robot to help you clean a kitchen?
3. Do you trust the robot to help you clean a bathroom?
4. Do you trust the robot to help you clean a window?
5. Do you trust the robot to help you with pest control?
6. Do you trust the robot to help you load dishes in a dishwasher?
7. Do you trust the robot to help you make a bed?
8. Do you trust the robot to help you clean the floor?
9. Do you trust the robot to help you take out the trash?
10. Do you trust the robot to help you do laundry?
Cognitive Tasks
11. Do you trust the robot to remind you to take medication?
12. Do you trust the robot to help you stay informed about the weather/news?
13. Do you trust the robot to help you learn how to use new technology?
14. Do you trust the robot to help you get information about your hobbies?

A.6 Acknowledgements

This material is based upon work supported by the Air Force Office of Scientific Research (AFOSR) through the award FA9550-18-1-0125, and National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. DGE-1650112. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or AFOSR.

I thank Auriel Washburn for her mentorship and stewardship on the project and Thomas An for his assistance with robot system development. This chapter contains material form ”Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust” by A. Washburn,A. Adeleye,T. An and L.D Riek, In Proceedings of ACM Transactions on Human Robot Interaction (THRI), 2020. This dissertation author was one of the primary investigators and authors of this work.

Bibliography

[1]

[2] Comprehensive categories across food and non-food products. https://www.rccgrandprix.ca/categories. Accessed: 2022-02-25.

[3] Ros wiki: rail_segmentation. http://wiki.ros.org/rail_segmentation. Accessed: 2022-02-25.

[4] Sales of robots for the service sector grew by 37% world wide. https://ifr.org/news/sales-of-robots-for-the-service-sector-grew-by-37-worldwide. Accessed: 2023-07-27.

[5] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[6] Nichola Abdo, Cyrill Stachniss, Luciano Spinello, and Wolfram Burgard. Robot, organize my shelves! tidying up objects by predicting user preferences. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[7] Akanimoh Adeleye, Jiaming Hu, and Henrik I Christensen. Putting away the groceries with precise semantic placements. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pages 2219–2224.

[8] Max Bajracharya, James Borders, Dan Helmick, Thomas Kollar, Michael Laskey, John Leichty, Jeremy Ma, Umashankar Nagarajan, Akiyoshi Ochiai, Josh Petersen, Krishna Shankar, Kevin Stone, and Yutaka Takaoka. A mobile manipulation system for one-shot teaching of complex tasks in homes, 2020.

[9] Dhruv Batra, Angel X. Chang, Sonia Chernova, Andrew J. Davison, Jia Deng, Vladlen Koltun, Sergey Levine, Jitendra Malik, Igor Mordatch, Roozbeh Mottaghi, Manolis Savva, and Hao Su. Rearrangement: A challenge for embodied ai, 2020.

[10] Jenay M Beer, Cory-Ann Smarr, Tiffany L Chen, Akanksha Prakash, Tracy L Mitzner, Charles C Kemp, and Wendy A Rogers. The domesticated robot: design guidelines for assisting older adults to age in place. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 335–342. ACM, 2012.

[11] Patrick Beeson and Barrett Ames. Trac-ik: An open-source library for improved solving of generic inverse kinematics. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 928–935. IEEE, 2015.

[12] Dmitry Berenson, Siddhartha S. Srinivasa, Dave Ferguson, and James J. Kuffner. Manipulation planning on constraint manifolds. In *2009 IEEE International Conference on Robotics and Automation*, pages 625–632, 2009.

[13] Vincent-Pierre Berges, Andrew Szot, Devendra Singh Chaplot, Aaron Gokaslan, Roozbeh Mottaghi, Dhruv Batra, and Eric Undersander. Galactic: Scaling end-to-end reinforcement learning for rearrangement at 100k steps-per-second. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13767–13777, June 2023.

[14] Antonio Bicchi and Vijay Kumar. Robotic grasping and contact: A review. In *Proceedings 2000 ICRA. Millennium conference. IEEE international conference on robotics and automation. Symposia proceedings (Cat. No. 00CH37065)*, volume 1, pages 348–353. IEEE, 2000.

[15] Rainer Bischoff. Design concept and realization of the humanoid service robot hermes. In *Field and Service Robotics*, pages 485–492. Springer, 1998.

[16] Elin A Björling and Narayan Singh. Exploring temporal patterns of stress in adolescent girls with headache. *Stress and Health*, 33(1):69–79, 2017.

[17] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[18] James Bruce and Manuela Veloso. Real-time randomized path planning for robot navigation. In *IEEE/RSJ international conference on intelligent robots and systems*, volume 3, pages 2383–2388. IEEE, 2002.

[19] Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. Experiences with an interactive museum tour-guide robot. *Artificial intelligence*, 114(1-2):3–55, 1999.

[20] Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Benchmarking in manipulation research: The ycb object and model set and benchmarking protocols. *arXiv preprint arXiv:1502.03143*, 2015.

[21] Sebastian Castro. Behavior Trees for Home Service Robotics Tasks. https://www.youtube.com/watch?v=xbvMnpwXNPk, 2022.

[22] Artificial Intellgence Center. Shakey the robot. 1984.

[23] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 541–548. IEEE, 2015.

[24] JYC Chen and PI Terrence. Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, 52(8):907–920, 2009.

[25] David Coleman, Ioan Sucan, Sachin Chitta, and Nikolaus Correll. Reducing the barrier to entry of complex robotic software: a moveit! case study. *arXiv preprint arXiv:1404.3785*, 2014.

[26] David Coleman, Ioan A. Şucan, Sachin Chitta, and Nikolaus Correll. Reducing the barrier to entry of complex robotic software: a moveit! case study. In *Journal of Software Engineering for Robotics*, page 5(1):3–16, 2014.

[27] Michele Colledanchise and Petter Ögren. *Behavior trees in robotics and AI: An introduction*. CRC Press, 2018.

[28] SN Copleston and Guido Bugmann. Personal robot user expectations. *Advances in Communications, Computing, Networks, and Security*, 5:230–238, 2008.

[29] Cynthia L Corritore, Beverly Kracher, and Susan Wiedenbeck. On-line trust: concepts, evolving themes, a model. *International journal of human-computer studies*, 58(6):737–758, 2003.

[30] Guowei Cui, Wei Shuai, and Xiaoping Chen. Semantic task planning for service robots in open worlds. *Future Internet*, 13(2):49, 2021.

[31] P Dario. Movaid: A mobile robotic system for residential care to disabled and elderly people. In *Proc. 1st MobiNet Symp., Athens*, pages 45–68, 1997.

[32] Paolo Dario, Eugenio Guglielmelli, Vincenzo Genovese, and Maurizio Toro. Robot assistants: Applications and evolution. *Robotics and autonomous systems*, 18(1-2):225–234, 1996.

[33] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. Impact of robot failures and feedback on real-time trust. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press, 2013.

[34] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of the AISB Convention: New frontiers in human-robot interaction*, pages 6–9, 2009.

[35] Sandra Devin and Rachid Alami. An implemented theory of mind to improve human-robot shared plans execution. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 319–326. IEEE, 2016.

[36] Carl F DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 321–326. ACM, 2002.

[37] Andreas Doumanoglou, Jan Stria, Georgia Peleka, Ioannis Mariolis, Vladimir Petrik, Andreas Kargakos, Libor Wagner, Václav Hlaváč, Tae-Kyun Kim, and Sotiris Malassiotis. Folding clothes autonomously: A complete pipeline. *IEEE Transactions on Robotics*, 32(6):1461–1478, 2016.

[38] Brian R Duffy. Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3-4):177–190, 2003.

[39] Matthew Dunbabin and Lino Marques. Robots for environmental monitoring: Significant advancements and applications. *IEEE Robotics & Automation Magazine*, 19(1):24–39, 2012.

[40] Mary T Dzindolet, Linda G Pierce, Hall P Beck, Lloyd A Dawe, and B Wayne Anderson. Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3):147, 2001.

[41] Peter F Edemekong, Deb L Bomgaars, Sukesh Sukumaran, and Shoshana B Levy. Activities of daily living. *StatPearls [internet]*, 2021.

[42] J. F. Engelberger. *Robotics in Service*. MIT Press, 1989.

[43] Kutluhan Erol, James A Hendler, and Dana S Nau. Semantics for hierarchical task-network planning. Technical report, 1995.

[44] Federal Reserve of Economic Data (FRED), St. Louis,. *E-Commerce Retail Sales*, Feb 2022.

[45] Richard E Fikes and Nils J Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4):189–208, 1971.

[46] Angel Flores-Abad, Ou Ma, Khanh Pham, and Steve Ulrich. A review of space robotics technologies for on-orbit servicing. *Progress in aerospace sciences*, 68:1–26, 2014.

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

[48] Jean Fox and Deborah Boehm-Davis. Effects of age and congestion information accuracy of advanced traveler information systems on user trust and compliance. *Transportation Research Record: Journal of the Transportation Research Board*, 1621:43–49, 1998.

[49] Sourav Garg, Niko Sünderhauf, Feras Dayoub, Douglas Morrison, Akansel Cosgun, Gustavo Carneiro, Qi Wu, Tat-Jun Chin, Ian Reid, Stephen Gould, Peter Corke, and Michael Milford. Semantics for robotic mapping, perception and interaction: A survey. *Foundations and Trends® in Robotics*, 8(1–2):1–224, 2020.

[50] Ashok K Goel, Tesca Fitzgerald, and Priyam Parashar. Analogy and metareasoning: Cognitive strategies for robot learning. In *Human-Machine shared contexts*, pages 23–44. Elsevier, 2020.

[51] Matthew Gombolay, Xi Jessie Yang, Bradley Hayes, Nicole Seo, Zixi Liu, Samir Wadhwania, Tania Yu, Neel Shah, Toni Golen, and Julie Shah. Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research*, 37(10):1300–1316, 2018.

[52] Birgit Graf, Matthias Hans, and Rolf D. Schraft. Care-O-bot II – Development of a next generation robotic home assistant. *Autonomous robots*, 2004.

[53] Birgit Graf, Matthias Hans, and Rolf D Schraft. Care-o-bot ii—development of a next generation robotic home assistant. *Autonomous robots*, 16(2):193–205, 2004.

[54] Victoria Groom and Clifford Nass. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, 8(3):483–500, 2007.

[55] Victoria Groom, Vasant Srinivasan, Cindy L Bethel, Robin Murphy, Lorin Dole, and Clifford Nass. Responses to robot social roles and social role framing. In *International Conference on Collaboration Technologies and Systems (CTS)*, pages 194–203. IEEE, 2011.

[56] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Open-vocabulary object detection via vision and language knowledge distillation. *arXiv preprint arXiv:2104.13921*, 2021.

[57] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G Pipe, and Kerstin Eder. Believing in bert: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 493–500. IEEE, 2016.

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

[59] Kerstin S Haring, Katsumi Watanabe, Mari Velonaki, Chad C Tossell, and Victor Finomore. Ffab–the form function attribution bias in human-robot interaction. *IEEE Transactions on Cognitive and Developmental Systems*, 2018.

[60] Cory J Hayes, Maryam Moosaei, and Laurel D Riek. Exploring implicit human responses to robot mistakes in a learning from demonstration task. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 246–252. IEEE, 2016.

[61] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[62] PM Herbers, DA Elder, and JG Woo. Area under a curve: calculation and visualization. In *Proceedings of the 2011 Midwest SAS Users Group. (Report No.: DG12–2011)*, 2011.

[63] Marion Hersh. Overcoming barriers and increasing independence–service robots for elderly and disabled people. *International Journal of Advanced Robotic Systems*, 12(8):114, 2015.

[64] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real-time loop closure in 2d lidar slam. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 1271–1278. IEEE, 2016.

[65] Laura M Hiatt, Anthony M Harrison, and J Gregory Trafton. Accommodating human variability in human-robot teams through theory of mind. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[66] Pamela J Hinds, Teresa L Roberts, and Hank Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181, 2004.

[67] Guy Hoffman and Cynthia Breazeal. Anticipatory perceptual simulation for human-robot joint practice: Theory and application study. In *AAAI*, pages 1357–1362, 2008.

[68] Shanee Sarah Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology*, 9:861, 2018.

[69] Kai-yuh Hsiao, Stefanie Tellex, Soroush Vosoughi, Rony Kubat, and Deb Roy. Object schemas for grounding language in a responsive robot. *Connection Science*, 20(4):253–276, 2008.

[70] Jiaming Hu, Akanimoh Adeleye, and Henrik I Christensen. Place-and-pick-based re-grasping using unstable placement. In *The International Symposium of Robotics Research*, pages 252–267. Springer, 2022.

[71] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning*, pages 9118–9147. PMLR, 2022.

[72] T. Iqbal and L. D. Riek. Human robot teaming: Approaches from joint action and dynamical systems. *Humanoid Robotics: A Reference*, 2017.

[73] Tariq Iqbal, Michael J Gonzales, and Laurel D Riek. Joint action perception to enable fluent human-robot teamwork. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 400–406. IEEE, 2015.

[74] Tariq Iqbal, Samantha Rack, and Laurel D Riek. Movement coordination in human–robot teams: a dynamical systems approach. *IEEE Transactions on Robotics*, 32(4):909–919, 2016.

[75] Alex Irpan, Alexander Herzog, Alexander Toshkov Toshev, Andy Zeng, Anthony Brohan, Brian Andrew Ichter, Byron David, Carolina Parada, Chelsea Finn, Clayton Tan, Diego Reyes, Dmitry Kalashnikov, Eric Victor Jang, Fei Xia, Jarek Liam Rettinghouse, Jasmine Chiehju Hsu, Jornell Lacanlale Quiambao, Julian Ibarz, Kanishka Rao, Karol Hausman, Keerthana Gopalakrishnan, Kuang-Huei Lee, Kyle Alan Jeffrey, Linda Luu, Mengyuan Yan, Michael Soogil Ahn, Nicolas Sievers, Nikhil J Joshi, Noah Brown, Omar Eduardo Escareno Cortes, Peng Xu, Peter Pastor Sampedro, Pierre Sermanet, Rosario Jauregui Ruano, Ryan Christopher Julian, Sally Augusta Jesmonth, Sergey Levine, Steve Xu, Ted Xiao, Vincent Olivier Vanhoucke, Yao Lu, Yevgen Chebotar, and Yuheng Kuang. Do as i can, not as i say: Grounding language in robotic affordances. 2022.

[76] Shrutheesh R Iyer, Anwesan Pal, Jiaming Hu, Akanimoh Adeleye, Aditya Aggarwal, and Henrik I Christensen. Household navigation and manipulation for everyday object rearrangement tasks. In *submision to 2023 7th IEEE International Conference on Robotic Computing (IRC)*.

[77] Advait Jain and Charles C Kemp. El-e: an assistive mobile manipulator that autonomously fetches objects from flat surfaces. *Autonomous Robots*, 2010.

[78] Keith S Jones and Elizabeth A Schmidlin. Human-robot interaction: toward usable personal service robots. *Reviews of Human Factors and Ergonomics*, 7(1):100–148, 2011.

[79] Yash Kant, Arun Ramachandran, Sriram Yenamandra, Igor Gilitschenski, Dhruv Batra, Andrew Szot, and Harsh Agrawal. Housekeep: Tidying virtual households using commonsense reasoning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, pages 355–373. Springer, 2022.

[80] Ivan Kapelyukh and Edward Johns. My house, my rules: Learning tidying preferences with graph neural networks. In *Conference on Robot Learning*, pages 740–749. PMLR, 2022.

[81] Lydia E Kavraki, Petr Svestka, J-C Latombe, and Mark H Overmars. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE transactions on Robotics and Automation*, 12(4):566–580, 1996.

[82] Gayane Kazhoyan, Simon Stelter, Franklin Kenghagho Kenfack, Sebastian Koralewski, and Michael Beetz. The robot household marathon experiment, 2021.

[83] Cory David Kidd. *Sociable robots: The role of presence and task in human-robot interaction*. PhD thesis, Massachusetts Institute of Technology, 2003.

[84] Ralf Kittmann, Tim Fröhlich, Johannes Schäfer, Ulrich Reiser, Florian Weißhardt, and Andreas Haug. Let me introduce myself: I am care-o-bot 4, a gentleman robot. *Mensch und computer 2015–proceedings*, 2015.

[85] Stefan Kohlbrecher, Johannes Meyer, Thorsten Graber, Karen Petersen, Uwe Klingauf, and Oskar von Stryk. Hector open source modules for autonomous mapping and navigation with rescue robots. In *Robot Soccer World Cup*, pages 624–631. Springer, 2013.

[86] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434, 2008.

[87] Alyssa Kubota, Emma IC Peterson, Vaishali Rajendren, Hadas Kress-Gazit, and Laurel D Riek. Jessie: Synthesizing social robot behaviors for personalized neurorehabilitation and beyond. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 121–130, 2020.

[88] Anthony R Lanfranco, Andres E Castellanos, Jaydev P Desai, and William C Meyers. Robotic surgery: a current perspective. *Annals of surgery*, 239(1):14, 2004.

[89] C Leclercq, P Allemand, A Balcerzak, F Branca, RF Sousa, A Lartey, M Lipp, VP Quadros, and P Verger. Fao/who gift (global individual food consumption data tool): a global repository for harmonised individual quantitative food consumption studies. *Proceedings of the Nutrition Society*, 2019.

[90] In Lee. Service robots: A systematic literature review. *Electronics*, 10(21):2658, 2021.

[91] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

[92] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. Gracefully mitigating breakdowns in robotic services. In *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 203–210. IEEE, 2010.

[93] John J Leonard and Hugh F Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *IROS*, volume 3, pages 1442–1447, 1991.

[94] Hector J Levesque. Knowledge representation and reasoning. *Annual review of computer science*, 1(1):255–287, 1986.

[95] Michael Lewis, Katia Sycara, and Phillip Walker. The role of trust in human-robot interaction. In Huseein A. Abbass, Jason Scholz, and Darryn J. Reid, editors, *Foundations of Trusted Autonomy*, pages 135–159. Springer-International, Cham, Switzerland, 2018.

[96] Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, et al. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. *arXiv preprint arXiv:2108.03272*, 2021.

[97] Ying Siu Liang, Damien Pellier, Humbert Fiorino, Sylvie Pesty, and Maya Cakmak. Simultaneous end-user programming of goals and actions for robotic shelf organization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[98] Weiyu Liu, Angel Daruna, and Sonia Chernova. Cage: Context-aware grasping engine, 2020.

[99] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg. Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. *arXiv preprint arXiv:1703.09312*, 2017.

[100] Matthew T Mason and J Kenneth Salisbury Jr. Robot hands and the mechanics of manipulation. 1985.

[101] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.

[102] Paul K McClure. "you're fired," says the robot: The rise of automation in the workplace, technophobes, and fears of unemployment. *Social Science Computer Review*, 36(2):139–156, 2018.

[103] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 2013.

[104] Andrew T Miller, Steffen Knoop, Henrik I Christensen, and Peter K Allen. Automatic grasp planning using shape primitives. In *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, volume 2, pages 1824–1829. IEEE, 2003.

[105] So Yeon Min, Devendra Singh Chaplot, Pradeep Ravikumar, Yonatan Bisk, and Ruslan Salakhutdinov. Film: Following instructions in language with modular methods. *arXiv preprint arXiv:2110.07342*, 2021.

[106] Matthias Minderer, Alexey Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, et al. Simple open-vocabulary object detection. In *European Conference on Computer Vision*, pages 728–755. Springer, 2022.

[107] Michael Montemerlo, Joelle Pineau, Nicholas Roy, Sebastian Thrun, and Vandi Verma. Experiences with a mobile robotic guide for the elderly. *AAAI/IAAI*, 2002:587–592, 2002.

[108] Masahiro Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

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

[110] Adithyavairavan Murali, Weiyu Liu, Kenneth Marino, Sonia Chernova, and Abhinav Gupta. Same object, different grasps: Data and semantic knowledge for task-oriented grasping, 2020.

[111] Dana Nau, Yue Cao, Amnon Lotem, and Hector Munoz-Avila. Shop: Simple hierarchical ordered planner. In *Proceedings of the 16th international joint conference on Artificial intelligence-Volume 2*, pages 968–973, 1999.

[112] Aastha Nigam and Laurel D Riek. Social context perception for mobile robots. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 3621–3627. IEEE, 2015.

[113] Cem Nizamoglu. Baghdad clock in aachen: Harun al rashid's gift to charlemagne, Dec 2021.

[114] Illah R Nourbakhsh, Judith Bobenage, Sebastien Grange, Ron Lutz, Roland Meyer, and Alvaro Soto. An affective mobile robot educator with a full-time job. *Artificial intelligence*, 114(1-2):95–124, 1999.

[115] IFR International Federation of Robotics. International federation of robotics.

[116] Sherry Ogreten, Stephanie Lackey, and Denise Nicholson. Recommended roles for uninhabited team members within mixed-initiative combat teams. In *Collaborative Technologies and Systems (CTS), 2010 International Symposium on*, pages 531–536. IEEE, 2010.

[117] Richard L Oliver. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, pages 460–469, 1980.

[118] Richard L Oliver and Gerald Linda. Effect of satisfaction and its antecedents on consumer preference and intention. *ACR North American Advances*, 1981.

[119] Weilyn Pa.

[120] Steffi Paepcke and Leila Takayama. Judging a bot by its cover: an experiment on expectation setting for personal robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 45–52. IEEE Press, 2010.

[121] Anwesan Pal, Yiding Qiu, and Henrik Christensen. Learning hierarchical relationships for object-goal navigation. In *Conference on Robot Learning*, pages 517–528. PMLR, 2021.

[122] Caroline Pantofaru, Leila Takayama, Tully Foote, and Bianca Soto. Exploring the role of robots in home organization. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 327–334, 2012.

[123] Eui Park, Quaneisha Jenkins, and Xiaochun Jiang. Measuring trust of human operators in new generation rescue robots. *Proceedings of the JFPS International Symposium on Fluid power*, 2008(7-2):489–492, 2008.

[124] David Paulius and Yu Sun. A survey of knowledge representation in service robotics. *Robotics and Autonomous Systems*, 118:13–30, 2019.

[125] Chris Paxton, Chris Xie, Tucker Hermans, and Dieter Fox. Predicting stable configurations for semantic placement of novel objects, 2022.

[126] Elizabeth Phillips, Scott Ososky, Janna Grove, and Florian Jentsch. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 1491–1495. SAGE Publications Sage CA: Los Angeles, CA, 2011.

[127] Aaron Powers and Sara Kiesler. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 218–225. ACM, 2006.

[128] Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. Virtualhome: Simulating household activities via programs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8494–8502, 2018.

[129] Lisa M PytlikZillig, Brittany Duncan, Sebastian Elbaum, and Carrick Detweiler. A drone by any other name: Purposes, end-user trustworthiness, and framing, but not terminology, affect public support for drones. *IEEE Technology and Society Magazine*, 37(1):80–91, 2018.

[130] Yiding Qiu, Anwesan Pal, and Henrik I. Christensen. Learning hierarchical relationships for object-goal navigation, 2020.

[131] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[132] Ulrich Reiser, Christian Connette, Jan Fischer, Jens Kubacki, Alexander Bubeck, Florian Weisshardt, Theo Jacobs, Christopher Parlitz, Martin Hägele, and Alexander Verl. Care-o-bot® 3-creating a product vision for service robot applications by integrating design and technology. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1992–1998. IEEE, 2009.

[133] Donna G Rhodes, Meghan E Adler, John C Clemens, and Alanna J Moshfegh. What we eat in america food categories and changes between survey cycles. *Journal of Food Composition and Analysis*, 2017.

[134] Florian Richter, Shihao Shen, Fei Liu, Jingbin Huang, Emily K Funk, Ryan K Orosco, and Michael C Yip. Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection. *IEEE Robotics and Automation Letters*, 6(2):1383–1390, 2021.

[135] Laurel Riek and Don Howard. A code of ethics for the human-robot interaction profession. *Proceedings of We Robot*, 2014.

[136] Laurel D Riek. The social co-robotics problem space: Six key challenges. *Robotics Challenges and Vision (RCV2013)*, 2013.

[137] Laurel D Riek. Healthcare robotics. *Communications of the ACM*, 60(11):68–78, 2017.

[138] Laurel D Riek, Woodrow Hartzog, Don A Howard, AJung Moon, and Ryan Calo. The emerging policy and ethics of human robot interaction. In *HRI (Extended Abstracts)*, pages 247–248, 2015.

[139] Elon Rimon. *Exact robot navigation using artificial potential functions*. Yale University, 1990.

[140] Paul Robinette, Ayanna M Howard, and Alan R Wagner. Timing is key for robot trust repair. In *International Conference on Social Robotics*, pages 574–583. Springer, 2015.

[141] Paul Robinette, Ayanna M Howard, and Alan R Wagner. Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, 47(4):425–436, 2017.

[142] Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404, 1998.

[143] Ericka Rovira, Austin Cross, Evan Leitch, and Craig Bonaceto. Displaying contextual information reduces the costs of imperfect decision automation in rapid retasking of isr assets. *Human factors*, 56(6):1036–1049, 2014.

[144] Deb Roy. Grounding words in perception and action: computational insights. *Trends in cognitive sciences*, 9(8):389–396, 2005.

[145] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems*, 2016.

[146] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–148. ACM, 2015.

[147] Jessica A Sandoval, Michael Ishida, Saurabh Jadhav, Sidney Huen, and Michael T Tolley. Tuning the morphology of suction discs to enable directional adhesion for locomotion in wet environments. *Soft Robotics*, 9(6):1083–1097, 2022.

[148] Ashutosh Saxena, Justin Driemeyer, and Andrew Y Ng. Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, 27(2):157–173, 2008.

[149] C Schaeffer and T May. Care-o-bot-a system for assisting elderly or disabled persons in home environments. *Assistive technology on the threshold of the new millenium*, 3, 1999.

[150] Martin J Schuster, Dominik Jain, Moritz Tenorth, and Michael Beetz. Learning organizational principles in human environments. In *2012 IEEE International Conference on Robotics and Automation*, pages 3867–3874. IEEE, 2012.

[151] Kerstin Severinson-Eklundh, Anders Green, and Helge Hüttenrauch. Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous systems*, 42(3-4):223–234, 2003.

[152] Dhruv Shah, Błażej Osiński, Sergey Levine, et al. Lm-nav: Robotic navigation with large pre-trained models of language, vision, and action. In *Conference on Robot Learning*, pages 492–504. PMLR, 2023.

[153] Saul Shiffman. Conceptualizing analyses of ecological momentary assessment data. *Nicotine & Tobacco Research*, 16(Suppl_2):S76–S87, 2013.

[154] Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. Alfred: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10740–10749, 2020.

[155] Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*, 2020.

[156] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.

[157] Matt Simon. 2017 was the year the robots really, truly arrived, Dec 2017.

[158] Randall Smith, Matthew Self, and Peter Cheeseman. Estimating uncertain spatial relationships in robotics. In *Autonomous robot vehicles*, pages 167–193. Springer, 1990.

[159] Harold Soh, Pan Shu, Min Chen, and David Hsu. The transfer of human trust in robot capabilities across tasks. *arXiv preprint arXiv:1807.01866*, 2018.

[160] Mohsen Soori, Behrooz Arezoo, and Roza Dastres. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 2023.

[161] Siddhartha S. Srinivasa, Dave Ferguson, Casey J. Helfrich, Dmitry Berenson, Alvaro Collet, Rosen Diankov, Garratt Gallagher, Geoffrey Hollinger, James Kuffner, and Michael Vande Weghe. Herb: A home exploring robotic butler. *Autonomous Robots*, 2010.

[162] Siddharth Srivastava, Shlomo Zilberstein, Abhishek Gupta, Pieter Abbeel, and Stuart Russell. Tractability of planning with loops. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[163] Simon Stepputtis, Joseph Campbell, Mariano Phielipp, Stefan Lee, Chitta Baral, and Heni Ben Amor. Language-conditioned imitation learning for robot manipulation tasks. *Advances in Neural Information Processing Systems*, 33:13139–13150, 2020.

[164] Yuan Sun. Effects of expectation setting and agency on human-robot interaction. Master's thesis, The Pennsylvania State University, State College, Pennsylvania, USA, 2011.

[165] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13438–13444. IEEE, 2021.

[166] Andrew Szot, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, Oleksandr Maksymets, et al. Habitat 2.0: Training home assistants to rearrange their habitat. *Advances in Neural Information Processing Systems*, 34:251–266, 2021.

[167] Akira Taniguchi, Shota Isobe, Lotfi El Hafi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Autonomous planning based on spatial concepts to tidy up home environments with service robots. *Advanced Robotics*, 35(8):471–489, 2021.

[168] Deborah Tannen. What's in a frame? surface evidence for underlying expectations. *Framing in discourse*, 14:56, 1993.

[169] Deborah Tannen and Cynthia Wallat. Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social psychology quarterly*, pages 205–216, 1987.

[170] Sebastian Thrun, Scott Thayer, William Whittaker, Christopher Baker, Wolfram Burgard, David Ferguson, Dirk Hahnel, D Montemerlo, Aaron Morris, Zachary Omohundro, et al. Autonomous exploration and mapping of abandoned mines. *IEEE Robotics & Automation Magazine*, 11(4):79–91, 2004.

[171] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[172] Ning Wang, David V Pynadath, and Susan G Hill. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 109–116. IEEE Press, 2016.

[173] Ning Wang, David V Pynadath, Ericka Rovira, Michael J Barnes, and Susan G Hill. Is it my looks? or something i said? the impact of explanations, embodiment, and expectations on trust and performance in human-robot teams. In *International Conference on Persuasive Technology*, pages 56–69. Springer, 2018.

[174] Shengye Wang and Henrik I Christensen. Tritonbot: First lessons learned from deployment of a long-term autonomy tour guide robot. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 158–165. IEEE, 2018.

[175] Samantha F Warta, Katelynn A Kapalo, Andrew Best, and Stephen M Fiore. Similarity, complementarity, and agency in hri: Theoretical issues in shifting the perception of robots

from tools to teammates. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, pages 1230–1234. SAGE Publications Sage CA: Los Angeles, CA, 2016.

[176] Auriel Washburn, Akanimoh Adeleye, Thomas An, and Laurel D Riek. Robot errors in proximate hri: how functionality framing affects perceived reliability and trust. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(3):1–21, 2020.

[177] David Watkins-Valls, Peter K Allen, Henrique Maia, Madhavan Seshadri, Jonathan Sanabria, Nicholas Waytowich, and Jacob Varley. Mobile manipulation leveraging multiple views, 2021.

[178] Katherine S Welfare, Matthew R Hallowell, Julie A Shah, and Laurel D Riek. Consider the human work experience when integrating robotics in the workplace. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 75–84. IEEE, 2019.

[179] Christopher D Wickens and Stephen R Dixon. The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3):201–212, 2007.

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

[181] Bohan Wu, Roberto Martin-Martin, and Li Fei-Fei. M-ember: Tackling long-horizon mobile manipulation via factorized domain transfer. *arXiv preprint arXiv:2305.13567*, 2023.

[182] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. Tidybot: Personalized robot assistance with large language models. *arXiv preprint arXiv:2305.05658*, 2023.

[183] Kimitoshi Yamazaki, Ryohei Ueda, Shunichi Nozawa, Yuto Mori, Toshiaki Maki, Naotaka Hatao, Kei Okada, and Masayuki Inaba. System integration of a daily assistive robot and its application to tidying and cleaning rooms. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.

[184] Naoki Yokoyama, Alexander William Clegg, Eric Undersander, Sehoon Ha, Dhruv Batra, and Akshara Rai. Adaptive skill coordination for robotic mobile manipulation. *arXiv preprint arXiv:2304.00410*, 2023.

[185] Georgios A Zachiotis, George Andrikopoulos, Randy Gornez, Keisuke Nakamura, and George Nikolakopoulos. A survey on the application trends of home service robotics. In *2018 IEEE international conference on Robotics and Biomimetics (ROBIO)*, pages 1999–2006. IEEE, 2018.

[186] Alireza Zareian, Kevin Dela Rosa, Derek Hao Hu, and Shih-Fu Chang. Open-vocabulary object detection using captions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14393–14402, 2021.

[187] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[188] Zhen Zeng, Adrian Röfer, and Odest Chadwicke Jenkins. Semantic linking maps for active visual object search, 2020.

[189] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *European Conference on Computer Vision*, pages 350–368. Springer, 2022.