Autonomy, Social Agency, and the Integration of Human and Robot Environments

by

Jack Thomas

M.Math, University of Waterloo, 2014
B.CS, University of New Brunswick, 2012 B.A., University of New Brunswick, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Science

SIMON FRASER UNIVERSITY
Fall 2019

Approval

Name:	**Jack Thomas**
Degree:	**Doctor of Philosophy (Computer Science)**
Title:	**Autonomy, Social Agency, and the Integration of Human and Robot Environments**
Examining Committee:	**Chair:** Nick Sumner Associate Professor
	Richard Vaughan Senior Supervisor Professor
	Mo Chen Supervisor Assistant Professor
	Parmit Chilana Internal Examiner Assistant Professor
	Bill Smart External Examiner Professor Collaborative Robotics and Intelligent Systems Institute Oregon State University
Date Defended:	**December 4th, 2019**

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

There is a growing wave of new autonomous robots poised for use in human environments, from the self-driving car to the delivery drone. By leaving the confinement of factories and warehouses for city streets and office buildings, the field of robotics is on a collision course with the wider public, and must prepare to contend with society as a whole's reaction to integrating human and robot spaces. In the research community, this is leading to a convergence of the fields of Autonomy and Human-Robot Interaction, producing new and emergent issues. This thesis proposes that one such issue is the problem of a robot's "Social Agency", whereby navigating among humans necessarily makes robotic agents part of society in the eyes of humans, and so robots must play a social role in order to achieve the acceptance they need to be effective.

After grounding this idea within existing theory, we will examine the Social Agency proposition over three parts, representing three research projects. In part one, we investigate the potential of an "incidental interface" for human-robot interaction that adapts an existing autonomous, multi-robot system to use audio for inter-robot communication, allowing human co-workers to supervise their work through casual overhearing. Part two re-implements another autonomous system for human-robot and robot-robot doorway navigation, where a user study finds a link between social acceptance of the robot, accepting the robot's right of way, and performance. Insights gained from that study are leveraged in part three with a redesigned doorway system that makes proactive, self-confident determinations about right of way, leading to the discovery of a polarizing reaction among participants dubbed "Agency Alienation". We close with an examination of what this development arc demonstrates about the potential and pitfalls of developing a robot's Social Agency, and what this may mean for the future of robotics in public spaces.

Keywords: Human-Robot Interaction; Autonomy; Social Agency

Acknowledgements

John Donne said that no man is an island, and while this thesis is presented as the culmination of my own research journey, none of it would have been possible alone.

I extend my gratitude and thanks to my supervisor, Professor Richard Vaughan, whose guidance, patience, and friendship led me through the five years of this degree. I am also thankful to Professors Mo Chen and Angelica Lim, each of whom served on my committee and provided much useful advice.

I am grateful to my labmates and other department colleagues, including Jacob Perron, Lingkang Zhang, Jake Bruce, Abbas Sadat, Mani Monajjemi, Shokoofeh Pourmehr, Jens Wawerla, Sepehr Pour, Geoff Nagy, Rakesh Shrestha, Payam Nikdel, Pratik Gujjar, Faraz Shamshirdar, Amirmasoud Ghasemi, and Bita Azari.

Many friends have also been there for me over the years. I owe a debt to Stephen Kiazyk, Oliver Trujillo, Laura Cang, Andrew Arnold, Aaron Moss, Cecylia Bocovich, Hella Hoffmann, Valerie Sugarman, Dean Shaft, David Dufour, and the rest of the crew from Waterloo. I am also thankful to the fine folks of Fredericton, such as Logan Marks, Amanda Lee Gallant, Alex Harris, Crystal Harris, Danny MacDonald, Taylor Seely-Wright, Brenden Roach, James Burnes, Chad Gerein, Mitchell MacLean, Brodie Beaman, Shawn MacKenzie, and Delia Dee. A few dear old friends deserve additional note, like Martin Jarman, Jared Morrison, and Dylan McGuire.

My girlfriend, Dom Corbett, is a constant source of inspiration, and her patient support during the writing of this thesis is warmly appreciated.

And, of course, I am grateful to my sister, Ceri Thomas, my father, Michael Thomas, and my late mother, Paula Thomas. Their love, support, and inspiration have been constants throughout my life, and I owe them everything I am today.

Many more people deserve recognition than time and space here will permit. I closed the acknowledgements of my Master's thesis with a cliche, “we stand on the shoulders of giants” - a classic passing of credit to the genius of those who came before.

The older I get, however, the more I see that these “giants” are not merely a few academic titans. They are a vast mountain of people, everyone around you, the whole community that in ways small and large lifts you up, bringing the most distant dreams within reach. What I have achieved was made possible only by the opportunities they gave me, and I humbly acknowledge their contribution.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A wave of new autonomous robots is poised to sweep into human environments. Where once robots were limited to controlled spaces in factories and laboratories, sequestered from the general public, now corporations and researchers aim to bring them to streets, skies, workplaces, and homes.

Leading technology and automobile companies are racing to bring the self-driving car to market[1]. Amazon's warehouses are some of the biggest integrated human-robot workplaces in the world, and they have toyed with the idea of UAV delivery drones[2] - while Ford and Agility Robotics have recently announced their own humanoid delivery robot[3]. Security patrol robots are being tested by companies like Knightscope[4] for malls, offices, and parking garages. SoftBank's humanoid robot, Pepper, has been tested as an advertiser and host outside restaurants in major American airports[5]. Until recently, the only real-life robot of this kind the average person was likely to encounter was the humble Roomba vacuum cleaner[6], but this era of public robotics as inoffensive curiosity is set to change.

Change is likewise coming to the academic study of robotics. This push into human spaces is also driving a convergence of the fields of autonomous robotics and human-robot interaction (HRI), one which challenges many of their base assumptions. Putting autonomous, multi-robot systems like the self-driving car into public streets necessitates losing control over the environment and the introduction of dynamic and unpredictable new agents - hu-

[1] http://fortune.com/2018/05/31/whos-winning-the-self-driving-car-race/

[2] https://www.nbcbayarea.com/news/local/Amazon-Delivery-Drones-Might-be-Coming-Soon-484789181.html

[3] https://www.cnn.com/2019/05/22/tech/ford-delivery-robot/index.html

[4] https://www.nanalyze.com/2018/05/robot-security-guards-knightscope/

[5] https://www.retailcustomerexperience.com/news/pepper-the-robot-now-an-airport-restaurant-greeter/

[6] https://www.zdnet.com/article/automation-how-irobots-roomba-vacuum-cleaner-became-part-of-the-family/

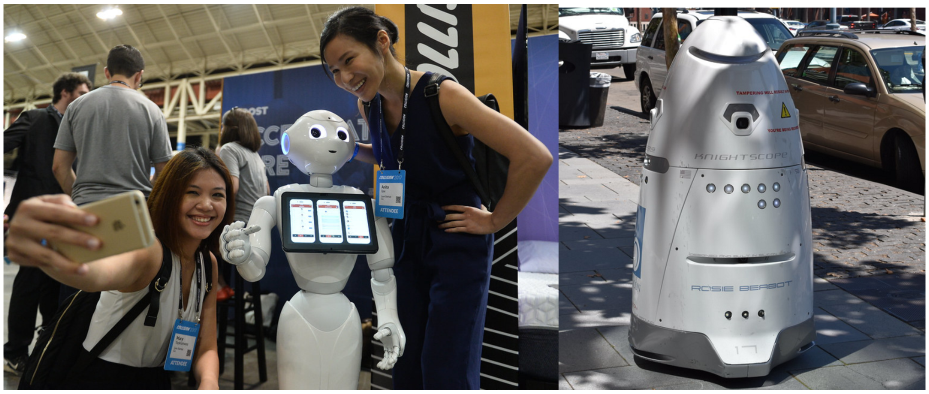

Figure 1.1: Autonomous robots for human environments include Softbank's Pepper and Knightscope's security patrol robot.

man beings. Conversely, keeping the old focus on 'Wizard of Oz' simulation experiments and theoretical speculation in HRI would overlook the opportunity to study humans and robots together in practice, in the wild, running live and goal-driven systems.

One point of disconnect for these two fields that may prove difficult to reconcile is their definitions of agent. Autonomy leans more toward the abstract entities of multi-agent theory (many multi-robot systems are effectively embodied multi-agent ones). This is partly informed by the meaning of autonomy in robotics, which Bekey gave as [9] "[referring] to systems capable of operating in the real-world environment without any form of external control for extended periods of time." By contrast, HRI draws from sociology and the humanities to define its agents through their social relations to other agents - robots as companions, interlocutors, even bystanders.

There exists a temptation that since each field articulates their systems and methods in terms of "agents", they may be mutually intelligible, and can be blended together. After all, the simplest way to integrate robots into human spaces that already have customs surrounding navigation and interaction would be to adapt autonomous methods that mimic these human behaviours and step seamlessly into the pre-existing navigational system. This minimizes the disruption for humans and gives robots access to new environments, a seemingly mutually-beneficial outcome.

Consider a self-driving car arriving at a four-way traffic stop. Both law and social custom hold that the first car to arrive at a four-way stop will have the right-of-way to pass through the intersection first, and a driver who goes out of turn is both violating the law and disrespecting their fellow driver. Now imagine a human driver arrives at an intersection only to find that a robot car (with no visible human passengers) has arrived first. The law may still hold that the robot be allowed to pass first, but if the human does not perceive the robot as a **Social Agent**, then there is no one to disrespect. Without the regulating effect of social judgement and in the absence of police threatening to enforce the law, why

Figure 1.2: Four-way traffic stops work by a combination of legal and social customs, but will these naturally extend to a robot driver?

not cut off that self-driving car? Is simply mimicking the expected social signals enough to earn the right of way for that car?

This risk of failure to recognize robotic **Social Agency** could hinder robots trying to navigate the social topography of public spaces. Many of the common customs of those spaces, like letting passengers disembark from a train first before boarding, or being able to ask for directions from passers-by in a crowded city, presume a degree of social recognition that goes deeper than sending the right outward signals. This is where the disconnect between the definitions of agent poses a challenge to the smooth integration of the two fields.

So what is Social Agency? On the HRI and Human-Computer Interaction side, social agency theory already has roots extending into Reeves and Nass's Media Equation [89] and Mayer's Social Agency Theory for Media Learning [67], exploring how computers can employ social cues to foster social responses and connections in users. In *When Artificial Social Agents Try to Persuade People: The Role of Social Agency on the Occurrence of Psychological Reactance* [93], by Roubroeks, Ham, and Midden, they give a definition that aligns well with our own usage: "Consistent with social agency theory, we use the term social agency to indicate the degree to which a social agent is perceived as a social entity. In other words, the degree to which a social agent is perceived as being capable of social behavior that resembles human-human interaction...". For our purposes, cultivating Social Agency in robots will mean gaining for them the social recognition of humans, and being

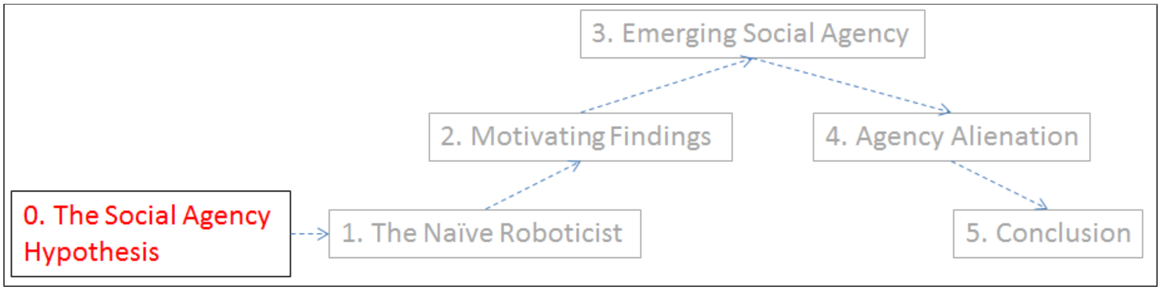

Figure 1.3: The Development Arc of Social Agency, Step 0: Introduction of the Social Agency hypothesis as the goal of the thesis and the development arc as one of two supporting rationales.

able to participate in reciprocal behaviours and interactions normally contingent on that recognition.

The overall hypothesis of this thesis, then, is **Social Agency represents a key emergent issue for the integration of human and robot environments**. Not simply that the concept exists, or that it has relevance, but that tackling Social Agency is crucial to solving the problem of bringing autonomous robots to the public - and if left unaddressed, may result in backlash that could bring down the entire project.

To test this hypothesis and its implications, we will present two different rationales, exploring both theoretical and experimental avenues. The first will be a **Theory of the Significance of Robot Social Agency**, which grounds the concept in the existing literature through academic references and inductive reasoning. The second is a three-part development arc of projects (previewed in Figure 1.3), each of which sought to adapt an autonomous, multi-robot behaviour for use with humans in new, shared environments. Starting by re-implementing a wifi-mediated task-allocation system into audio before switching to resolving navigational deadlocks around doorways, these projects gradually refine and explore the idea that the main obstacle to their success in the wild is their acceptance by the humans whose world they have intruded upon. Each project includes a study with regular participants, whose natural reactions, perceptions, and testimony form the basis by which our own theory of Social Agency is both developed and evaluated.

The contributions of this thesis are divided into two categories:

1) The overall investigation of Social Agency. This includes further defining Social Agency and evaluating its' significance to the creation of integrated human-robot environments, as well as identifying implications of this significance for roboticists and society at large. This is achieved both through an examination of existing social robotics theory, and an iterative development process of redesigning autonomous robot systems for use in human environments then evaluating them through user studies.

2) The individual contributions toward robot autonomy in human environments made by each of the three projects: i) The "incidental interface" audio approach for human co-workers

to supervise multi-robot systems. ii) The “assertive” doorway negotiation behaviour for solving human-robot and robot-robot deadlocks around doors. iii) The redesigned “assured” doorway negotiation behaviour.

Structurally, this introduction will be followed by a related work chapter that revisits the prior depth report’s investigation of the Autonomy/HRI overlap and provides additional grounding of the Social Agency concept. Chapters three, four, and five will cover the three projects that provide the core contents of this thesis (the incidental interface, the assertive doorway behaviour, and the assured doorway behaviour) drawn from the original three research papers they appeared in. Finally, Chapter six will assess what the findings of those projects say about the necessity of Social Agency research, as well as offer comment on the entire human-robot integration endeavour the field is now undertaking.

Chapter 2

Related Work

The first chapter laid out an ambitious, high-level sketch about the state of robotics, the merging of two research fields, and the somewhat abstract phenomenon of robotic Social Agency. Before we proceed to testing this idea experimental, we first need to contextualize our claims within the existing literature.

To that end, this chapter will present the independent theoretical justification for Social Agency in robotics that complements the experimentally-derived development arc. These two approaches will independently support our assertion that Social Agency is an emergent issue worthy of closer study. We will also take a step back to consider larger philosophical discourse surrounding robotics and AI, and then return to autonomy and HRI research to find supporting examples of their convergence.

2.1 Social Agency

While Social Agency may not yet have a significant presence in robotics, it has some history already in human-computer interaction research. One example of this is *When Artificial Social Agents Try to Persuade People: The Role of Social Agency on the Occurrence of Psychological Reactance* [93], by Roubroeks, Ham, and Midden. Their experiment explicitly modelled the impact of Social Agency on the perceptions of their study participants, by varying the level of embodiment they gave to the instructions read to participants - either as plain text (low agency), dialogue from a still image of a robotic face (mid agency), or dialogue from a video clip of a talking robotic face (high agency). The study found some evidence to suggest that receiving controlling instructions from increasingly agent-like sources would create higher degrees of negative push-back, a “reactance” to the assertions of an artificial social agent that will become relevant to our work again in later chapters.

While the idea of Social Agency may already be present elsewhere, our decision to articulate it within HRI did not spring up arbitrarily. The depth report that preceded this thesis developed a five-point theory regarding the significance of Social Agency to the current moment in robotics, laid out in Table 2.1..

Table 2.1: Theory of the Significance of Robot Social Agency

1. There is a desire for autonomous robots that can operate in human environments.
2. In order to be autonomous, these robots will have to be intelligent and independent agents.
3. Humans perceive their own agency as different from robots.
4. Social Agency requires the mutual recognition of and respect for each other's intentionality.
5. Therefore, the advancement of long-term robot autonomy in human environments will eventually require developing robot Social Agency

The first clause of the theory, **‘There is a desire for autonomous robots that can operate in human environments,"** has already been put forward in the previous chapter, which described a number of autonomous robot projects and products. This desire is what motivates us to ask what it would take for autonomous robots to succeed in said human environments, and so take a closer look at what it means to be autonomous.

2.1.1 Further Defining Autonomy

In *Autonomy in Robots and Other Agents* [102], Smithers criticizes what he sees as the vague and inconsistent use of the term autonomy in computer science and robotics when compared to other fields, such as biology, law, philosophy and so on. In these areas, autonomy is tied strongly to the concept of self-identity and "self-law-making", a layer of actualization that goes beyond the mere self-regulating controllers found in existing robotic systems.

This disconnect between theoretical definitions of autonomy also plays out within the field of practical research. In *Elephants Don't Play Chess* [18], Brooks notes that the thrust of AI research concerned with logical reasoning as the path to true general intelligence ignores that the development of such intelligence in living beings did not run through this route - in essence, that while a computer might beat an elephant at chess, we would consider the elephant much closer to us in terms of intelligence.

Why is an elephant's type of intelligence more recognizably autonomous to us than a computer's? In Pfeifer's *Building “Fungus Eaters”: Design Principles of Autonomous Agents* [85], self-sufficiency and the means to maintain it are proposed as the basis of autonomy for an agent - a fungus-eating robot stranded on a distant world must have the means to seek out and eat fungus in order to survive. Similarly, an elephant's intelligence allows it to feed itself and maintain its independence, while a chess-playing computer's intelligence does nothing for its' survival.

This centering of self-sufficiency gives us our second clause, **“In order to be autonomous, these robots will have to be intelligent and independent agents,”**
.

2.1.2 The Distinction Between Human and Robot Autonomy

If the bar for recognizable autonomy is only self-sufficiency, is socially integrating autonomous robots with humans as simple as making self-sufficient robots? In *When are robots intelligent autonomous agents?* [103], Steels argues that many more academic standards of intelligence used in the field are insufficiently persuasive for general use. In describing agents, he notes the notion of self-interest and self-maintenance as motivation for an agent's actions is key - a human travel agent, for example, does their work in return for compensation. They have a sense of personal agency that extends beyond the boundaries of their immediate role. Steels concludes that by his more exacting standard, no robot has yet become intelligent or autonomous.

This distinction between robots and humans need not be a simple categorical one. In *Mixing human and non-humans together: The sociology of a door closer* [50], Johnson notes that sociology does not have to limit itself to social relations between humans, or even between living beings - any entity that can hold up any sort of relation to another can be a site of sociological study, even a humble door-closing mechanism. Certainly the human tendency to anthropomorphize an unreliable car or a fond old tool should be familiar to most. In *A Cyborg Manifesto* [44], Haraway expounds at length that the 20th century had breached all conceptual barriers between man and animal, or man and machine, spreading the notion that we are all just physical constructs of varying complexity.

This is where the disconnect between the autonomy and HRI definitions of agent can be found. In *Ants Don't Have Friends - Thoughts on Socially Intelligent Agents*, Dautenhan [24] describes a category difference between "artificial" agents, who exist only within the confines of a scenario (a factory robot that ceases to exist as an independent agent when the factory is closed), and "organic" agents, who can take on a role but also have lives outside of any one scenario (an employee who can take their hat off at the end of the day and become a regular person once again). This distinction has nothing to do with the substance of the agent - an organic agent could be made of metal, an artificial one flesh and blood - only the nature of their existence. This gives us the third clause, **"Humans perceive their own agency as different from robots."**

Conveniently, we already have a working example of what this distinction and tension between artificial and organic agents looks like - as Kuipers explores in *An Existing Ecologically Successful Genus of Collectively Intelligent Artificial Creatures* [59], businesses have already undergone this sort of social integration. They possess legal recognition, they feed in order to survive, they interact both with each other and with individual humans, yet despite all these similarities, there is an immensely significant distinction in how people perceive their place in society compared to a fellow human. Former Presidential candidate Mitt Romney said "Corporations are people,", but they are neither citizens nor friends.

2.1.3 The Missing Piece in Social Recognition

These qualms give us the fourth clause, **"Social Agency requires the mutual recognition of and respect for each other's intentionality."** As Breazeal argues in *How to Build Robots that Make Friends and Influence People* [16], in order to maintain a social relationship to another agent, humans need a sense of their "intentionality" - the sense that the other party has internal beliefs, that they act with intent, in essence that they themselves are also able to maintain a social relationship.

Without this sense of mutual recognition, a human can at best only pantomime respecting the social rights of the other party, as they cannot sustain a belief in them. We can see the end of such illusions when discussing robots as moral agents, which force the issue of the lack of perceived independence. In *Android arete: Toward a virtue ethic for computational agents* [21], Coleman deliberately sidesteps the problem of artificial agents lacking their own will by articulating what "is very consciously a slave ethic". Robots may act with moral competence, or contribute to a moral outcome, but we cannot yet see them as personally moral, revealing our true feelings that they do not share equally in the responsibilities (and thus recognition) of society.

The fifth and final clause - the guiding principle of the report's subsequent review of Autonomy and HRI literature - was the conclusion that **"Therefore, the advancement of long-term robot autonomy in human environments will eventually require developing robot Social Agency"**. This assertion stands as the natural consequence of the four preceding premises, for if the belief in a robot's Social Agency is a precondition for recognition of their agency by humans, robots will be unable to integrate socially in human society, and so limit their potential as autonomous agents.

While this theory may provide a clean rationale for investigating Social Agency, it also tells only half the story of how this thesis came about. At the beginning of the series of projects described in the coming chapters, the notion of Social Agency was not yet the motivating force. We began with only the first premise, that autonomous robots in human environments were a rising phenomenon, which motivated an investigation into the obstacles and opportunities they presented. It was the evidence produced by studies and experimentation that prompted searching the literature for supporting theories to organize these results into a coherent narrative.

Therefore, it is worth keeping in mind that the route this thesis will take toward discovering the significance of Social Agency will follow a much more roundabout and organic process, relative to the theory presented in Table 2.1, which will involve developing methods, evaluating them with the public, and searching for a common cause to explain their reactions. In this way, both the theoretical justification presented here and the arc of our development of autonomous systems for human spaces will lead us toward the same conclusion.

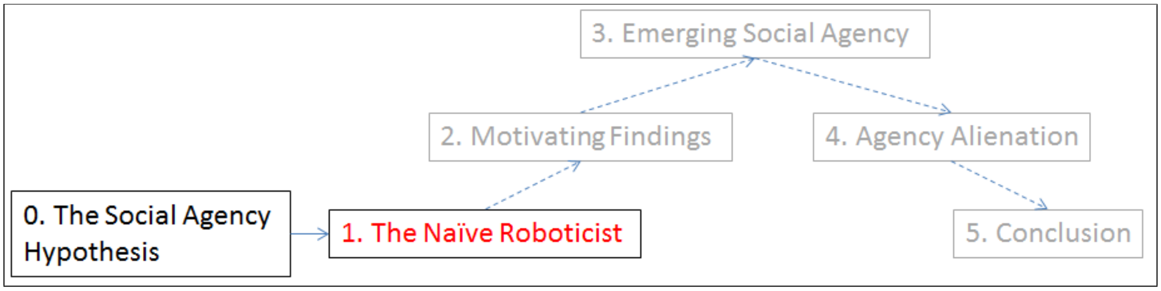

Figure 2.1: The Development Arc of Social Agency, Step 1: Beginning from the perspective of a roboticist seeking to bring autonomous robots to human environments.

2.2 Philosophy of Technology

It is worth taking an aside to integrate our more narrow focus on the concept of Social Agency into the larger philosophical discourse surrounding artificial intelligence and robotics. Raising the notion of an artificial agent claiming a social right cannot help but bring to mind the debates surrounding the fundamental nature of AI itself.

The classical centerpiece of AI theory is Turing's *Computing Machinery and Intelligence* [108], which gave us the famous "Turing Test". This proposed that if a human communicating via a computer terminal with both another human and a computer couldn't tell which was which, meaning that the computer could perfectly imitate a human, then the distinction of which is "real" no longer matters - from the perspective of the human, they are both equally real. By this argument, if our artificial social agents can mimic all the outward signals and behaviours of a sociable human being, they have an equal claim to being accepted into human society.

Opposed to this viewpoint is Searle's Chinese Room Argument [98], which recasts the situation of the Turing Test as a man who does not speak Chinese being locked in a room with a book of instructions, receiving messages from the outside and sending corresponding messages back out. No matter how sophisticated the book of instructions is, and how persuaded the people outside the room might be that they are communicating with someone inside, the trapped man still does not speak Chinese and cannot be said to understand the conversation he facilitates - and similarly, a computer cannot have semantic understanding of its actions. Under this view, artificial Social Agency will remain all but impossible - and attempts to achieve it are misguided at best, actively deceptive at worst.

Regardless of the answer to this intractable debate, others have noted some observable phenomena regarding general attitude and reaction to artificial agents. Reeves and Nass, for example, put forward what is known as the Media Equation [89], that humans interacting with computer-based media cannot help but trigger social receptors. They also found, in *Can Computer Personalities be Human Personalities* [79], that personality traits (in this case, dominance vs. submission) can be expressed via computers through interactions with seemingly no overt social framing, yet nevertheless provoke some recognition and affinity from persons of a matching personality. These separate the question of whether one *can* create a convincing imitation of a social agent, from whether one *should*.

Higher philosophical questions have bearing on our original motivation of smoothing the integration of autonomous robots into human environments. As Bonnefon et al. note in *The Trolley, The Bull Bar, and Why Engineers Should Care About The Ethics of Autonomous Cars* [12], the arrival of the self-driving car has already triggered far-reaching and fundamental moral debates relevant to their regulation by governments and acceptance by society. The "Trolley Problem", normally a contrived introduction for philosophy students to moral dilemmas, is almost exactly recreated by the circumstances of a runaway self-driving car deciding which group of pedestrians to barrel into.

This word of warning is meant for those who consider only the engineering performance aspect of the problem faced by autonomous robots in human environments. The higher questions provoked by the creation of artificial Social Agents cannot be neatly siloed off as a purely academic curiosity, and when it comes time to start regulating or even voting on the matter of robotics' expanding reach into society, it may no be only the philosophers asking these questions.

2.3 The Overlap of Autonomy and HRI

Another major premise of our project is that the advent of autonomous robots coming to human environments is driving the fields of robot autonomy and human-robot interaction together. Since this is presented as the cause of Social Agency's new urgency, examining the specific points of relation between the two will provide useful context for the specific projects described in the coming chapters. It also affords us the opportunity to acknowledge where other researchers are already beginning to adapt to this shifting ground.

2.3.1 Recharging

The ability of a robot to keep itself charged is a bedrock capability for autonomy - Pfeifer's fungus-eaters would be nothing without the ability to eat fungus. The now-defunct robotics research lab Willow Garage made a point of developing this capability for their own robots, as in *Long Term Autonomy in Office Environments* [70], which explored how long their PR2 robot could keep itself charged without human intervention (13 days, it would turn out). Other considerations for the self-charging robot include navigating the obstacles of an indoor environment that might stand between them and their charging dock. To that end, *Autonomous Door Opening and Plugging in With a Personal Robot* [71] investigated how to traverse office doorways and plug into outlets.

As an autonomous activity, recharging contains a whole host of potential costs, but also opportunities. In *Staying Alive Longer: Autonomous Robot Recharging Put To The*

Test [101], the adoption of a recharging mechanism is presented as a precondition for true long-term autonomy, and immediately turns their research robot into a 24-hour sentry for the lab. In *Basic cycles, utility and opportunism in self-sufficient robots* [68], this intuition is codified into an evaluative framework to assess the time and energy cost efficiency of different behaviours with different robots.

This expansion of recharging considerations to the surrounding environment naturally shades into social dimensions in HRI. In *Exploring socially intelligent recharge behaviour for human-robot interaction* [26], a study found users were more forgiving of a robot stopping a helpful behaviour to recharge itself if the robot excused itself verbally first. Earlier, in *Managing social constraints on recharge behaviour for robot companions using memory* [27], a robot was set to learn patterns of user behaviour throughout the day and choose periods of historically low traffic to recharge itself - resulting in less inconvenience to users and less chance of disruption for the robot. Recharging may at first have appeared as a purely autonomous behaviour, but as these works show, once robots are expected to manage their energy in public spaces, they begin facing the same sort of social expectations and etiquette that humans have surrounding their own self-maintenance.

2.3.2 Scheduling, Task Allocation, and Authority

Once a robot is charged and operational, it must be put to work. Another defining feature of autonomy is reducing the need for direct human supervision by having robots manage their own schedules and allocating themselves as needed. In *A Comprehensive Taxonomy for Multi-Robot Task Allocation* [57], Korsah argues most forms of work one might want to divide among a team of robots can be categorized according to four types of problem complexity, proposing generalized solutions that affect broad swaths of autonomous operation.

Some approaches to allocation solve the problem globally over all robots and distribute the solution, in the manner of a team coming together to work out a schedule. In *Fair subdivision of multi-robot tasks* [47], for example, a central server connecting a heterogeneous group of robots works out solutions to match the available capabilities of different robots to the needs of different types of work.

This allocation becomes more difficult when dealing with distributed autonomous robots without central control, requiring creative individual policies to generate the desired overall outcome. For example, in *A Fast and Frugal Method for Team-Task Allocation in a Multi-Robot Transportation System* [113], robots carrying supplies from sources to sinks can distribute themselves according to the productivity of each source by implementing a simple measure of patience - the longer the wait, the more likely a robot is to give up and try another queue, eventually reaching equilibrium. This is comparable to how customers allocate themselves at supermarket checkout lanes, gravitating toward short and fast-moving queues to naturally increase the system's throughput without the need for direct supervision or communication.

Once autonomous robots begin working alongside humans, however, this sort of distributed and independent decision-making means putting robots in a position to exercise some authority over their co-workers' requests. For example, in *Socially-Driven Collective Path-Planning for Robot Missions* [48], the authors tackle the idea of a waypoint-navigating robot where multiple users may each nominate waypoints for the robot to visit. However, with only limited time and energy available, a robot may not be able to complete all of these instructions - how then should it decide which waypoints to skip? The authors explore different approaches to fairness toward the different users, all of which hinge on the notion that the robot can be empowered to accept some orders and disregard others, and that spurned users will defer to the robot's judgement the way they might to a fellow human.

This question of equal partnership between robots and humans to achieve efficient task allocation leads directly into topics of HRI inquiry. In *Decision-Making Authority, Team Efficiency and Human Worker Satisfaction in Mixed Human-Robot Teams* [39], the authors created an experiment where a participant worked with both other humans and a robot to fetch and assemble part kits, with the goal of assembling them efficiently and several tasks to allocate. Their expectation was that humans would prefer a mixed responsibility for task allocation between themselves and the robot, but participants were willing to cede total allocation control to the robot once they understood it would improve their team's efficiency. Here we can see one of the hints at Social Agency caused by bringing HRI and Autonomy together, for while humans can be convinced to work with robots where they perceive a joint interest, this is not the same as acknowledging the robot has some manner of right.

Human coworkers' willingness to recognize robotic authority may be quick to fade if something goes wrong. In *When the robot criticizes you...: Self-serving bias in human-robot interaction* [119], participants were asked to recreate the motion that a robot would demonstrate, and the robot would then evaluate their performance - secretly, whether the evaluation was positive, neutral or negative was chosen at random, rather than actually depending on their performance. As expected, those participants who received negative evaluations were significantly more likely to shift the blame onto the robot, with lower views of its functionality or sociability. The authors note that this self-serving bias is visible when humans are taught by other humans, of course, but a robot tutor adds a whole new dimension of credibility for affronted humans to cast doubt on.

That was not the only project to find evidence humans look more fondly on robots that agree with them. In *Critic, compatriot or chump? Response to robot blame attribution* [41], participants tried to teach a humanoid robot their preferences among a set of pictures of objects, and when the robot would deliberately fail, the robot would either blame themselves, the team, or the participant. Unsurprisingly, the participants reacted negatively to being blamed, seeing the robot not just as less friendly or social, but less competent as well. Similarly, in *How a robot should give advice* [106], participants were shown videos of a robot or human assistant offering advice to someone baking cupcakes, with different videos varying the directions to be more commanding or include more hedging and "discourse markers" (broadly, informal language). Phrasing the instructions as casual advice over direct commands was better received with both human and robot assistants.

Does this mean that humans only appreciate autonomy in robots when it is not exercised against them? In *Evidence that Robots Trigger a Cheating Detector in Humans* [64], a study involving a robot cheating in a game of cards produced an interesting result. In those rounds where the robot cheated in order to win the game, users were more likely to describe the action as deliberate and intentional, while if the robot cheated in the player's favor or at random, this was seen as faulty behaviour. As we considered while laying down our own theory, acting in accordance with some perceived self-interest is recognizable to humans as agentic behaviour.

Robots that lack this recognition may be at a disadvantage when trying to apply social influence. In *Eyewitnesses are misled by human but not robot interviewers* [10], participants were shown a video of a crime in progress and then asked by either a robot or a human interviewer about the events of the video. Both human and robot interlocutors read from the same script, but participants were significantly more likely to be misled by leading questions from humans than from robots. The levers of peer pressure that could force a witness to change their story proved less accessible to robots.

This lack of social access by robots can influence their ability to carry out tasks with a social dimension. In *Exploring Minimal Nonverbal Interruption in HRI* [97], concerns imported from Human-Computer Interaction about the toll taken on users by interruptions informed their design of the most minimal non-verbal set of actions a robot could take to signal their need to interrupt. This approach implicitly suggests that the preferable way for a robot to assert itself is as lightly as possible.

Interruption is not inherently an act of antagonism, however. In *Designing interruptive behaviours of a public environmental monitoring robot* [29], the duties of an environmental monitoring robot might require interrupting humans in a possibly hazardous area to ask them about suspicious smells. Even though the robot's purpose is benign, the urgency of a pollutant leak may demand assertive action take precedence over social niceties. In their early study results, researchers found that when soliciting bystanders, politeness provoked more cooperation than empathy. A robot with duties to perform might not be rude, but it may perhaps be firm.

2.3.3 Navigation for Humans and Robots

Navigation is another core component of autonomy, and one that we will engage with significantly in the coming chapters. For the roboticist, navigation begins with purely theoretical algorithms, solving in abstract the questions of path planning and obstacle avoidance. The A* algorithm [45] is not specific to robotics - it is a graph traversal algorithm with many other applications in computer science - but it is an example of one such generalized method imported by robot planners.

Transitioning from pure theory to practical, real-world navigation requires addressing the specific needs of robots - for example, the Dynamic Window Approach (DWA), articulated in *The Dynamic Window Approach to Collision Avoidance* [33]. This method incorporates the physical constraints of a robot in motion, such as turn rates, speed and acceleration, to produce trajectories better adapted to real conditions. The approach has found traction among robot enthusiasts, coming prepackaged in the popular open-source robot software ROS's [87] standard navigation stack.

Even more challenging is considering the motion of other agents, which is where robot navigation begins dealing more closely with the specific needs of autonomy. In *Motion Planning in Dynamic Environments Using Velocity Obstacles* [32], the positions and velocities of other agents in the environment - velocity obstacles - are considered when making the robot's own plans about where to go to ensure they won't instigate a collision.

Some of the most sophisticated of these navigation methods will incorporate the dynamic obstacles' own agency, predicting how they will react to the robot's presence. *Reciprocal velocity Obstacles for real-time multi-agent navigation* [110] extends the velocity obstacle to consider each agent's mutual interest in avoiding collisions while maintaining efficiency, allowing agents to depend on one another to smooth their trajectories.

The intuition that cooperation makes for mutually satisfactory outcomes is easy to fulfill when the controllers of every agent are specified by the developer. In *A Decentralized Approach to Formation Maneuvers* [61], for example, groups of robots can be made to adopt different useful formations without the need for direct control - robots are moving in a distributed, autonomous manner, but the overall result can be directed.

This kind of cooperative, autonomous navigation is a natural candidate for the sort of naive transition from controlled robot spaces to uncontrolled human ones that was cautioned against earlier. After all, if reciprocal velocity obstacles works with autonomous navigation agents, shouldn't it extend naturally to human agents? To test this, *Local reactive robot navigation: A comparison between reciprocal velocity obstacle variants and human-like behavior* [43] weighed RVO approaches against human-inspired navigation methods in simulation. Considering safety and throughput, the authors found a human-like approach often outperformed the more abstract competitors.

Pursuing explicitly human-like and human-compliant navigation lead to many projects that capitalized on what we know about humans in particular over autonomous agents more generally. In *A navigation system for robots operating in crowded urban environments* [60], the robot's navigation mission covers kilometers of city ground, so the researchers embrace "pedestrian-like" behaviour as the goal of their system, recognizing the boundaries of sidewalks, traffic lights, and other urban navigation customs. Meanwhile, in *Dynamic path planning adopting human navigation strategies for a domestic mobile robot* [120], knowledge about human strategies for navigating their own home is leveraged to improve robot performance at the same task. In both cases, the route to a successful robot navigation controller runs through learning from how humans already navigate these environments.

Studying human environments in this way necessitates considering how to interact with not just individuals, but great masses of people. In *PLEdestrians: A Least-Effort Approach to Crowd Simulation* [42], thousands of simulated agents recreate the natural dynamics of humans moving as a crowd, including features like lane formation and crowd compression. In further work from the same lab, *Directing Crowd Simulations Using Navigation Fields* [83], this crowd simulation was upgraded to include field-based control mechanisms to shape the flow and direction of the simulated mass of human-like agents. The testing opportunities for simulated robot navigation controllers to interact with masses of humans without the need for large groups of volunteers are obvious, but the simulated crowd agents themselves can be a source of insight for how to blend in with a group.

In other cases, particularly swarm robotics, inspiration on group dynamics can also come from the wider animal kingdom. Reynolds' Boids, a software simulation detailed in *Flocks, herds and schools: A distributed behavioral model* [90], is a widely-cited model for recreating the movements of flocks of birds, herds of beasts and schools of fish - or potentially robots meant to simulate these things. For example, in *Self-organized flocking in mobile robot swarms* [107], where the tiny Kobots are made to travel in a group as one "super-organism", Reynolds' flocking model is cited as the origin for this brand of robot behaviour.

As for what it takes to make robot navigation more natural and comfortable for human interlocutors, the study of proxemics has thoroughly explored the sort of speeds and distances robots should maintain. In *Evaluation of Passing Distance for Social Robots* [80], for example, a user study gauging the range of comfortable distances for a robot to maintain while passing a human in the hall made the expected observation that passing too close was displeasing to participants. However, it also found that passing too far - especially if done sharply, such as a robot at the far end of the hall abruptly hugging the wall - was similarly awkward.

Relative positioning can also be thought of as a communication signal. In *Toward Understanding Social Cues and Signals in Human-Robot Interaction: Effects of Robot Gaze and Proxemic Behaviour* [31], distance was found to be more effective at signalling intent than visual gaze for human subjects passing in a corridor. That repeated interactions were found to solidify this signal also has implications for robot behaviour in the long-term, promoting consistency.

The distances and positioning modelled in proxemics have deeper implications than, say, the ideal emergency stop distance for a given robot obstacle avoidance controller. In *Human-Robot Proxemics: Physical and psychological distancing in human-robot interaction* [75], a study found a relationship between the likeability of the robot and the proximity the human felt comfortable maintaining. - and that this relationship was two-way, where an unwelcome gaze could drive a participant further away.

Gathering these sorts of human-factor insights allows for the development of human-like models of navigation, such as in *Social Forces Model for Pedestrian Dynamics* [46], a hand-crafted model for explicitly describing the motion of humans in public spaces. Robot controllers built on this model, such as the one described in *Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model* [100], find their success is enhanced beyond simple collision-avoidance approaches in pure navigation, as the more socially-compliant robot slots into human environments with a minimum of disruption.

Alternatively to model-building, machine learning offers new and increasingly competitive methods to capture the subtleties of human-compliant navigation. In *Social LSTM: human trajectory prediction in crowded spaces* [2], we see a machine-learning based example of the previous crowd behaviour simulations that were based on hand-crafted rule-sets. Using videos of crowds and neural network techniques, the resultant system can predict the trajectories of pedestrians on video with comparable accuracy to the social forces model yet with none of the explicit, laborious modelling.

This is not meant to imply that machine learning is blind to insight or domain knowledge. In *Learning social etiquette: human trajectory understanding* [91], significant thought is put into the construction and labeling of a dataset in order to highlight different types of mobile agents (such as bicyclists). Afterward, the variance in different humans' 'social sensitivity' lead to the positing of different navigation styles, discerning between a thoughtless rush and a casual stroll.

Machine learning approaches can also be used to recreate behaviours particular models are meant to capture. In *Socially compliant mobile robot navigation via inverse reinforcement learning* [58], human compliance is associated with participating in cooperative navigation behaviours, the same intuition behind RVO, and with enough data demonstrating this sort of reciprocation between humans, robots can be taught to mimic and participate in this human activity as well.

Techniques for learning from humans can also take inspiration from how humans learn. In *Socially Aware Motion Planning with Deep Reinforcement Learning* [20], it proved easier to train a model through punishing violations of certain navigation customs (e.g. passing to the left rather than the right) rather than try to positively define how to navigate. This sort of trial-and-error approach to learning makes intuitive sense to us when imagining ourselves in the robot's place, which can bolster our confidence in the result when they appear to take away the correct lesson.

What distinguishes these methods from machine-learning-based systems under general navigation is broadening the scope of learning to incorporate the same sort of concerns a human would have beyond simple point-to-point route planning. In *Socially Adaptive Path*

Planning in Human Environments Using Inverse Reinforcement Learning [54], maximal path efficiency is eschewed in favor of finding the most human-like trajectory - a goal all the more important since the robot being tested is an autonomous wheelchair, with a human passenger who would naturally like their navigation to be as seamless and smooth among fellow humans as if they were walking. Recalling Haraway's cyborgs [44] from our earlier background reading, we can already see some blurring of the line between human and robot agency in a robot learning to move with a human's social priority.

Whether by hand-crafted models or machine-learning, the insights gleaned concerning autonomous navigation around humans have already begun leading researchers into adapting their methods. In *Viewing Robot Navigation in Human Environments as a Cooperative behaviour* [53], their planner for robot motion is built from the ground up to consider the trajectories of both robots and humans as a joint-action problem, where the robot is concerned for the wellbeing and success of both agents on their respective journeys. Certainly a socially responsible mindset, but could it be naively optimistic to assume this reflects human social reality?

Even simply deciding where to stand relative to a human can carry significant social implications. In *Understanding suitable locations for waiting* [56], the proposed system is simply concerned with how to stay out of the way while a human they attend on is shopping, classifying potential waiting locations by how they could interfere with shop activities or pedestrian travel - not only could a badly-chosen spot displease a passing human, it might reflect socially poorly on the attended human for their robot to be so insensitive. In *How to Approach Humans? Strategies for Social Robots to Initiate Interactions* [96], lessons learned from an earlier, simpler system for approaching people in public led to an upgraded system that acknowledged the need to perform the right nonverbal social signals for attention before launching into a spoken question.

Positioning and motion has proven to be a powerful social signal to human observers. In *Perception of Affect Elicited by Robot Motion* [95], varying acceleration and trajectory curvature alone was enough to consistently communicate an internal 'emotional' state for the robot, regardless of the limitations of embodiment. Motion can even be used to convey fairly subtle ideas, such as deception - in *An Analysis of Deceptive Robot Motion* [28], a robot arm could use its motion alone to feint toward one of two water bottles, tricking a human participant in order to grab the unguarded one.

Nonverbal communication proves to be a crucial part of the social signalling built around human navigation, and many studies and systems have sought to explore it. In *Nonverbal leakage in robots: communication of intentions through seemingly unintentional behaviour* [77], two humanoid robots - one human-like in appearance, the other more stylized and exaggerated - played a game with human participants where they would secretly pick one of a set of objects, and the human would try to guess which one they had chosen by asking yes or no questions. By using slight, seemingly unintentional gaze cues toward their chosen object, both robots were able to let slip to the human which object they had chosen. While not an explicit part of the game, humans could naturally interpret the slippage, as though the robot was having a similar experience that a human in the same role might.

These sorts of non-verbal signals among humans go beyond the passive to the active, directing movements as much as any spoken instruction. In *Making a case for spatial prompting in Human-Robot Communication* [40], the researchers noted how in many human-robot interactions, humans would reflexively use hand gestures or body posture to indicate a stop or encourage the robot to take a certain position, even when recognizing these gestures was not incorporated into the interaction system. As for the other direction, in *Nonverbal robot-group interaction using an imitated gaze cut* [55], a mix of physical motion and gaze cues were used by a robot to single out a person from a crowd in order to deliver a parcel to them, the sort of delivery activity one might expect from commercial robots in the future.

Navigation presents a rich environment for exploring the intersection of the needs of both autonomy and HRI. The desire to solve pure engineering considerations for the sake of efficiency brought in by one field can conflict with or complement the desire to create a socially smooth experience by the other, and there are many opportunities for thoughtful adaptation. It is for these reasons that navigation will feature prominently in the research presented in the coming chapters, and may prove an important area for Social Agency research in the future.

2.4 Social Human-Robot Interaction and Human Factors

Our desire to encourage social sensitivity when moving autonomous robots into human environments is not entirely novel - Social Human-Robot Interaction is already an established subject where many of the topics we have broached are discussed. In *Social Interactions in HRI: The Robot View* [15], Breazeal stakes out social HRI as addressing the relationship between a human and a robot not as one to a tool or object, but to a creature or partner. Unlike in Human-Computer Interaction, Breazeal suggests, the designer must consider both the human and the robot's perspective during development.

To help us adopt this perspective, this section provides a broad survey of topics from Social HRI as well as Human Factors research. While not all of these works may be directly applicable to the projects we will be presenting in later chapters, the concepts and vocabulary they introduce help build the necessary context that our research is taking place within.

2.4.1 Trust

One way to measure a person's overall confidence in another is trust. Humans have a tendency to invest trust even in inanimate objects or abstract ideas, and robots are no exception to this, but lost trust will almost guarantee a loss of compliance by a human for a robot’s autonomy, or even a whole field of robotics - hence the justified fear of autonomous car researchers of the damage even one accidental death might have.

In *Trust-Driven Interactive Visual Navigation for Autonomous Robots* [116], a flying robot explicitly models the trust they predict their user has in their autonomous navigation based on how frequently the human sees fit to intervene by taking direct control, and adjusts their degree of autonomous control over their flight accordingly. Later, in *Maintaining efficient collaboration with trust-seeking robots* [117], they attempt to recognize when a user has lost faith in an autonomous car system by taking direct control of the vehicle and trying to convince the user they have fixed the error through a graphical interface, explicitly asking for the user to release control and give the autonomous vehicle another chance.

Recognizing the interpersonal relationship aspects of trust, the need to recognize when trust is lost, and the social cues surrounding asking for trust to be restored makes it easier for a human to put their faith in a robot (or the robot’s developers) by recreating the familiar routine of trust-based interactions with other humans. Once again, the more the robot portrays itself as a human-like social agent with human-like responsibilities, the easier it becomes for the human to reciprocate and release control to the robot’s autonomy.

2.4.2 Robots Asking for Help From Humans

Both recharging and navigation are examples of tasks where the help of humans might sometimes be necessary, particularly to mitigate an unexpected failure. In the aforementioned Willow Garage study of long-term office survival [70], the robot only failed when it fumbled its charging cable, the sort of minor inconvenience that asking a passing human to help with could easily resolve - so long as the human is willing to assist.

In *Robots asking for directions - The Willingness of passers-by to support robots* [114], an autonomous, mobile robot was tasked with finding its way to a goal location in a city with no prior map, instead asking bystanders to point the way and tap a few directions into a touch screen. Despite no incentive to help, the researchers found the robot could reliably reach the goal location with this incidental assistance, based on nothing more than the common courtesy humans might extend one another while out and about.

Similarly, in *Socially Distributed Perception* [72], a robot released at a robotics conference engages in games of "social tag", where they must track down a particular human by asking bystanders if they had seen them based on a physical description. While a conference-going audience may be more receptive than average at supporting a robot’s requests, they nevertheless found interesting variations in receptiveness depending on the social purpose of the space (the reception hall vs. a corridor) or the use of an ‘engaging’ movement by the robot, such that the robot was more successful when it conformed to human social customs. The authors introduce the idea of "socially distributed perception", articulating that so long as robots are limited in their sensing and actions, developers should explicitly seek to augment them with the help of friendly humans, especially when solving a difficult perception problem can be avoided with a simple social interaction.

Both projects imply accepting robot integration into human society (and conscious acceptance by humans) as a necessary condition for autonomy - after all, even the most independent-minded of humans depends on those around them from time to time, and a genuinely autonomous robot must not neglect the opportunity to do the same.

2.4.3 Social Integration of Robots

This is not the first time the issue of robot social integration has been raised, considering some of the most basic activities we might expect from an autonomous robot could hinge on this acceptance. In *A social robot that stands in line* [78], most of the work goes into the engineering for a robot that can recognize a queue for service (such as buying a coffee), find the end of the queue and wait at an appropriate place to join it, but this effort would be wasted if the next person to come along decided not to recognize the robot's right to claim a place in line. Over 20 trials, the researchers noted six failures relating to sensor issues or unexpected behaviours, and in these cases the robot would be ignored, but so long as the robot functioned as intended then customers would dutifully queue up behind it.

Integration goes beyond a single interaction, however - we must also consider that a long-term autonomous robot will have repeated interactions, often with the same people, and that attempting to build a true relationship with others would impose a whole new array of challenges. In *Designing robots for long-term social interaction* [38], a receptionist robot spent nine months installed in the entrance of a University building, providing basic services like directions and weather forecasts but also injecting some personality into these interactions, even an evolving backstory that expanded over time, in an effort to build a rapport with users. The authors found that while many users continued to use 'Valerie', few interactions lasted longer than 30 seconds.

Part of the puzzle of long-term acceptance may be that relationships are typically bi-directional. In *Personalization in HRI: A Longitudinal Field Experiment* [62], researchers had a snack delivery robot engage in standard interactions with half of the participants and personalized, adaptive ones with the other half - often as simple as apologizing for past mistakes or noticing patterns in past orders. Those participants who experienced these personalized interactions had increased rapport, engagement, and ultimately cooperation with the robot.

This connection between a robot's practical and personal interactions with humans has been observed elsewhere. In *Social vs. Useful HRI: Experiencing the Familiar, Perceiving the Robot as a Sociable Partner and Responding to Its Actions* [4], a humanoid NAO robot engaged in both social (greeting/goodbye) and 'useful' (handing over a survey) HRI, with particular attention paid to how human participant behaviour in one form of activity related to the other. While participants were more willing to participate in the useful interaction than the social ones overall, participation in social interactions also correlated with higher participation in the practical ones.

In *Human-Robot Teams Collaborating Socially, Organizationally and Culturally* [30], the authors identify three levels of social context a robot must contend with - the immediate team, the organization the team is a part of, and the wider culture containing both. This sense of social responsibilities outside of the immediate bounds of the interaction is a major step toward considering the robot a social agent unto itself, and recalls Dautenhan's definitions of artificial and organic agents by giving new and higher contexts for robots to exist within.

Perhaps the most direct case for the social integration of robots and humans by an HRI project can be found in *Toward Sociable Robots* [14]. Breazeal's robot, Kismet, is explicitly presented as a "social creature", a class of robot distinct from ones that are simply socially evocative, receptive, or interfacing - that is, it is not a robot to be used as a tool where its sociability is just a capacity, but exists as an entity unto themselves. This social creature status is in line with Breazeal's theories on intentionality, and the system presented where Kismet engages in verbal turn-taking with humans, expressing itself and learning from the expressions of others, is a case study of what robotics under a socially autonomous paradigm might look like.

2.4.4 Influence of Perception

Perception can strongly shape what social features humans attribute to objects, including possibly agency. In *Beyond Dirty, Dangerous and Dull: What Everyday People Think Robots Should Do* [104], a major survey was undertaken to find what jobs people thought robots suitable for. According to that survey, memory, perception, and service are valued in robots, while attributes like artistry, evaluation, judgment and diplomacy are reserved for humans. The authors note that this is actually a step up for robots, beyond being relegated to menial and repetitive machine functions. There may be some space for robots as competent and autonomous servants, but the reservation of "higher functions" for humans suggests there could still be resistance to the concept of robot agency.

Nevertheless, there may be means to overcome this perception. In *Robots that express emotion elicit better human teaching* [63], humans were asked to demonstrate dances for a robot to emulate. The robot was then scored on its performance, and the human was offered the opportunity to demonstrate the dance again for the robot. Those robots that reacted with appropriate emotion to their evaluation would consistently provoke humans to continue teaching them, and more accurately, than those who reacted inappropriately or not at all. Engaging human empathy with a robot's reactions and behaviours can have an impact on the final outcome of their interaction.

The stakes of this perception can also run quite high. In *Daisy, Daisy, give me your answer do!: Switching off a robot* [7], participants cooperated with a robot to play the classic board game Mastermind, with some combination of high or low agreeability and intelligence on the part of the robot determining the quality of its advice and the politeness of its suggestions. After the interaction was complete, participants were asked to turn off the robot over its own spoken objections, and participants with the more intelligent and agreeable robot would take up to three times longer to switch the robot off. Framed in those terms, meddling with what people might interpret as matters of life and death just to build a marginally more useful tool can take on significant moral hazard for an unwary designer. If anything, this paper treads dangerously close to the infamous Milgram experiment on obedience [73], where participants were deceived into thinking they were electrocuting an unseen person.

If we are right about the significance of social agency to autonomous robots, and that influencing perception is a means to achieving it, the consequences could extend far beyond what we have considered so far. In *A peer pressure experiment: Recreation of the Asch conformity experiment with robots* [13], the authors recreate an experiment where a group of either human test conductors or robots and one participant would be presented with the visual task of ordering lines from shortest to longest or the verbal task of pronouncing the past tense of a given set of words. In both tasks and with both the human assistants and the robots, some error would be made, and the measure of conformity was whether the participant would go along with the error. The authors found conformity to be significantly higher with humans than with robots, where robots had hardly higher influence than completing the task alone. The way in which humans distinguish between other humans and robots may - for now - protect them from being socially influenced by robots, but if this perception changes, the power to promote conformity may be in roboticists' hands.

2.4.5 Significance of Appearance

As a sub-topic of perception, appearance is one of the strongest signals for a robot's social standing, just as it is among humans. Apart from functional demands about whether a robot has the hardware to complete its tasks, our concern regarding a robot's appearance is expectations are set by this first impression. In *Judging a Bot By Its Cover: An Experiment on Expectation Setting for Personal Robots* [81], participants were presented with one of two different entertainment robots (the Pleo and the AIBO) while test conductors would either lower or raise expectations of its ability before allowing them to interact with the robot. The results showed participants were susceptible to expectation-setting, such that a robot exceeding low expectations was more highly thought-of than the same robot just meeting high expectations, and disappointing high expectations is unsurprisingly acutely negative - even if the same robot is used throughout. In brief, first impressions matter, and the danger of disappointing the often high bar set for robot functionality is real.

First impressions are not just constrained to the performance of the individual robot, either. In *Rabble of Robots Effects: Number and Type of Robots Modulates Attitudes, Emo-*

tions, and Stereotypes [35], participants were shown videos of either single or groups of robots, with either human-like, animal-like, or machine-like bodies, interacting with humans. They found that the interaction between these two variables had a more significant impact on participant impressions than either variable in isolation - for example, that a pack of machine-like iRobot Creates provoked much stronger negative reactions than a single Create, yet a group of human-like NAOs were more encouraging than a single NAO. One explanation the authors suggest for this reaction is that a group context will highlight a robot's social attributes, making the human more aware of their relation to the robot as a class rather than an individual. This can work to the advantage of a robot designed with socialization in mind, like the NAO, while emphasizing the more alien features of others, like the Create.

Knowing that different features of a robot's appearance and context can interact when humans form their impression necessitates a larger model for understanding this process. In *The Advisor Robot: Tracing People's Mental Model From a Robot's Physical Attributes* [86], researchers examined how changing the head dimensions and vocal features of a robot meant to give medical advice would impact user cooperation with that advice. From the responses of the participants, such as finding a masculine voice more 'knowledgeable' to associating a more human-like face with higher sociability, the authors deduced that humans start building a mental model for every new robot they meet and hunt for familiar cues they can use to infer the robot's properties.

It would be seriously reductive to take this model-building process to mean that the key to social recognition is simply to be as human-like as possible. In *Affective Expression in Appearance-constrained Robots* [11], four non-anthropomorphic robotics projects are assessed for how they made use of expressive modalities like body movement, posture, orientation, colour and sound to signal proxemic information to humans, allowing even a radio-controlled tank to engage socially. This should come as no surprise, from R2-D2 to WALL-E audiences have been charmed by the seeming humanity of the least human-like actors on film.

These concerns about expectations and satisfaction might be enough to justify why a roboticist should take their robot's appearance seriously, but the implications can actually run much deeper, as seen in *Which Robot Am I Thinking About?: The Impact of Action and Appearance on People's Evaluations of a Moral Robot* [66]. Asked to assess the "trolley problem", a classic ethics case study about whether it is ethical to throw a switch that will cause a runaway trolley to kill one person in order to save four others, participants predominantly believed the human was committing a moral hazard by killing one to save four, but the robot was simply being logical. However, when the robot presented alongside the problem was represented as physically more human-like, participants would correspondingly attribute more human-like moral responsibility to it, and recognize greater moral hazard in acting lethally.

This finding suggests a possible link between recognizing the agency of the robot and its moral culpability. In *Do people hold a humanoid robot morally accountable for the harm it causes?* [51], participants engaged with Robovie, a humanoid robot, by playing a visual scavenger hunt game with them. At the end of this interaction, the robot would incorrectly assess that the participant failed the game and deny them a monetary prize. In post-experiment interviews, two thirds of participants attributed moral responsibility to Robovie itself for having wronged them - less so than they might a human, but moreso than they might a vending machine. The success of the researchers at fueling the participants belief in Robovie as a social agent throughout the interaction with its behaviours also fueled their belief that Robovie held moral responsibilities.

Finally to bridge us to our next topic, in *Three's company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA* [34], researchers compared the reactions of participant pools drawn from two different countries to groups vs. individual robots and social vs. functional robots. Curiously, while participants from both countries preferred single social robots or group functional robots, Japanese participants looked at and interacted with robots they did not like more than those they liked, while American participants did the opposite. These differences between the participant pools add a whole new dimension to the concerns raised about expectation-setting, and will complicate the task of development considerably.

2.4.6 Influence of Background

We should close this chapter with a word of warning about the danger of generalizing too freely from any of these studies. Culture, demographics, personal history and a thousand other details influence the interactions between people, and we could reasonably assume the same of their interactions with robots. Any one study will face practical limitations on its sample of participants, but some researchers have explored just how significant these limitations are on results.

The first variation to be wary of is familiarity with the subject matter. In *On the Effect of the user's background on communicating grasping commands* [88], the distinction being considered is between technical and non-technical users when using verbal commands to direct a robot arm to grasp a series of small objects. As one might expect, the untrained users used more and simpler commands than trained lab members had, taking longer to complete the same tasks. This result is a reminder of the frequent danger studies and evaluations face in a lab-based setting, where the nearest pool of participants (university students) represents an abnormally expert population.

Different cultural backgrounds and norms can also impact a population's comfort with robots. In *When in Rome: the role of culture and context in adherence to robot recommendations* [112], a study compared U.S. and Chinese participants with robot collaborators who provided advice either explicitly or implicitly, from the belief that the Americans would respond better to the former and the Chinese to the latter, according to different standards of etiquette. As expected, each participant group saw greater cooperation with their robot when it exhibited a matching behaviour.

This evidence of variation need not be taken only as a blow to the validity of the studies we have covered in this report. In fact the most promising finding, when considering social recognition for robotic agents, may be that human attitudes appear to be flexible. Where researchers have found resistance to robotic integration today, there may be grounds for cooperation tomorrow. Whether this flexibility could also prove to be a danger, if the public is given reason to be suspicious of social integration with robots, is something we will revisit in later chapters.

Chapter 3

Humans Supervising Robot Co-workers Through the Incidental Interface

The pursuit of Social Agency begins, both in theory and experimental development, with the desire to integrate the next generation of autonomous robots into human environments. Our first investigation, then, started by looking at the needs of an autonomous robot in a mixed human-robot workplace, and what new developments may be needed to both ensure and expand on that autonomy.

The field of Human-Robot Interaction is typically concerned with direct and deliberate interactions, such as dialog systems and learning from demonstration. By contrast, when everything is going well in a mixed workplace, humans and robots may only need to interact occasionally as each gets on with their individual roles. Amazon's warehouses have only partially integrated humans and robots, but still represent a good example of two pools of workers active alongside one another and primarily interacting amongst themselves.

The trend in industrial workplaces is likely to continue towards large robot populations and relatively few workers, with Amazon again providing the example through their Picking Challenge [22] aiming to shrink the staff of their warehouses. This motivated our interest in blending HRI methods and autonomous systems for large-scale multi-human, many-multi-robot teams.

This project begins with a distributed multi-robot task allocation system we have already introduced in Chapter two, *the Fast and Frugal Sustain and Resupply System* (FASR) [113]. Originally conceived of as a purely autonomous robotics project, under FASR the robots communicate amongst themselves to achieve the optimal allocation of robots to the currently available tasks.

Our proposal is to redesign this system for use in human environments by adding the 'incidental interface'. We modify this multi-robot transportation system so that robots communicate with each other by audio signal instead of the default (silent) network messages.

Previously, the choice of communication medium made no difference to the efficacy of FASR, but making this component audible now allows for incidental supervision of the robots by humans. Through a user study, we show that untrained participants playing the role of co-workers are able to improve their situational awareness of the robots and resolve problems more quickly through overhearing these audio messages.

The incidental interface essentially provides a free HRI benefit through reassessing an autonomous system's implementation. This is an example of a proposed 'dual use' design principle, where exploring implementation options for other autonomous systems may also uncover HRI opportunities. Turning back to our original motivation, with many autonomous robots soon coming to human environments, many more autonomous systems could gain from such an assessment.

The idea to leverage ambient sound as an interface medium builds on the human-factor research of O'Shea et al. [36], who used the simulated sounds of a cola factory to promote cooperation between two human co-workers. This concept transitions fairly naturally from human-human to human-robot interaction.

To evaluate the incidental interface, we conducted a study to see if a simulated audio implementation could communicate to untrained humans some aspect of the dynamic state of a multi-robot system, such that the humans could decide when to intervene to assist the robots. We compare the system performance between audible and silent robots, where the performance of the robots in either case without human intervention would be identical. We also gathered user feedback on which implementation participants preferred and whether the sounds produced to manage robot allocation were sufficiently interpretable.

The individual contributions of this project are:

- Proposing and evaluating the incidental interface, an audio-based implementation for a task allocation system, where incidental sound can benefit the situational awareness of humans and improve throughput.

- Articulating the concept of dual use design, whereby autonomous systems being redesigned for use in human environments should be assessed for opportunities to achieve dual autonomy and HRI goals, as with the incidental interface.

- The feedback from the user study which, apart from providing useful evaluating responses concerning the incidental interface, will also reveal a sense of separation between the participants and their simulated co-workers that will go on to motivate future projects.

3.1 Related Work

The work described in this chapter touches on existing research in task allocation and human-robot interaction, both topics we introduced in Chapter two. We must also examine how audio has been used so far in robotics and compare it to progress made in the larger interfaces field.

3.1.1 Multi-Robot Task Allocation

Task allocation is a major topic of study within autonomous robotics, where a wide variety of methods exist to address the constraints of different sorts of problems. This includes the transportation task, which Gerkey's taxonony of task-allocation problems [37] would classify as a 'single-task robot / multi-robot task / instantaneous assignment' problem, meaning that each robot is engaged in one task at a time while each task can take on multiple robots, and allocation decisions must be made with whatever information is at hand in the moment. This problem family is also known as 'coalition formation'.

Coalition formation is a well-established topic in task allocation literature. Shehory and Kraus [99] describe a variety of algorithmic approaches. When it comes to implementation, however, the system imagined by Shehory and Kraus is an all-software network of agents, handling information and computing tasks. The subset of the field dealing with multi-robot systems, as surveyed by Yan et al.[118], often presents methods as theory rather than material systems. This leaves implementation details and their attendant opportunities underexplored.

3.1.2 Human Robot Interaction

In large-scale multi-robot systems, users may need a gestalt understanding of the state of the whole system, or large sub-population. Lopes et al.[65] addressed this issue when discussing how to supervise and control robotic swarms. Their experiments with the Kilobots made use of their coloured LED and light sensor to propagate states like group assignment through the swarm via coloured light, allowing human operators to assess the situation with a glance. That the lights are used for both communication between robots and to human supervisors makes it a strong example of a dual-use implementation.

This question of interfacing with large multi-robot groups was considered in Daily et al.[23], which proposed using augmented reality to show arrows anchored to each Pherobot robot pointing the way to an intruder as a means of turning the whole group into a collective visual system. This system required the user to wear a dedicated head-mounted receiver and display.

Leveraging human workplace cooperation to help robots solve difficult problems has some precedent. In Rosenthal and Veloso[92], a cooperative robot provides a human with information and guidance around a building but relies on the human to assist it in return like opening doors. This idea of cultivating the desire to help the robot in human partners through mutually-beneficial interactions is one we will return to.

3.1.3 Human Factors

The usefulness of audio in interfaces for situational awareness is already well-studied in human factor and HCI literature. One such example is earcons (“ear icons”) by Greenberg et al.[1], which have explored communicating complex information through principled sound design. Further research by Edwards et al.[17] established that well-designed earcons had a measurable impact on the performance of untrained human participants.

A notable example we have already raised as inspiration for this work is by O’Shea et al.[36], where a simulated cola-bottling factory is used to study how sound can be used to improve the collaboration between two human co-workers. When comparing audio-based and silent versions of the same experiment, they found that sounds generated by workers transmitted useful information to co-workers. Not only was each participant’s performance at controlling their half of the factory improved, but also cooperation was promoted between the humans through improved situational awareness of each other’s status. Creating this sense of co-worker collaboration between humans and robots would be a significant gain for human-robot interaction.

Adding sound to an interface is not guaranteed to benefit an interaction, however. Carlis et al.[5] found in their study that interrupting a user engaged in another task carried a cost both in performance on their own work and the user’s annoyance. Audio that is poorly implemented could fail to improve or even have a negative impact on the interaction experience.

3.1.4 Audio in Robotics

Audio as a tool in robotics has been arguably underutilized, particularly non-verbal audio that stands apart from work in natural language. Our experiment is inspired by the wonderfully minimal and elegant ‘chorusing’ concept from Holland et al.[49], which showed that robot team memberships could be controlled using only simple, repeated audio chirps to communicate. At the other extreme, in Cha et al.[19], robots that spoke while completing a physical task were perceived as more capable than silent robots, a result the authors attributed to the natural human tendency to anthropomorphize. This tendency is not limited to things which speak, meaning even non-verbal audio may be able to take advantage of this phenomenon.

As a prerequisite to reproducing some human and animal audio abilities, a major area of research is sound-source localization. Deleforge and Horaud[25], for example, shows a robot at a cocktail party separating and identifying different speakers. Pioneering work in this area includes Valin et al.[109], which took on the implementation challenges involved in bringing real-time sound-source separation to robots, followed by Michaud et al.[3] which evaluated the suitability of different audio localization algorithms for use with robots.

Closely related to this project, McLurkin et al.[69], tackled the topic of interfacing between humans and robot swarms directly through light, sound and behaviour. The potential of audio as a collective output channel is grounded by McLurkin in what he terms “nostalgia”, referring to veteran software engineers who debugged programs by listening to their computers. McLurkin also briefly raises the idea of using musical theory to connect to users. The purpose of this paper is to make explicit the idea of the incidental interface: we make decisions when building our system that will cause their behaviour to transmit information to nearby humans.

3.2 System

In order to demonstrate our dual-use design concept, we chose to extend FASR[113], a previous project on multi-robot task allocation. FASR is a distributed approach to balancing the number of robots engaged in different tasks such as transportation, and therefore is a useful example of existing multi-robot systems which could be ported to an integrated human-robot workplace. Like many other systems it also leaves open the question of exact implementation details.

3.2.1 FASR

FASR is a controller for individual, autonomous robots engaged in a transportation task. In this task, sources are producing resources that robots must collect and deposit at sinks. The number of robots each source needs to maximize throughput is a function of that source’s output rate, but this rate is unknown and may fluctuate. The challenge is allocating the right number of robots to each source as quickly as possible with no global information or central control. A Stage[111] simulation of a FASR system is shown in Figure 3.3, where small red robots are collecting yellow resource pucks from the light green and blue source squares and taking them back to the darker sink squares. The robots are optimally allocated to the available tasks. For clarity, here we consider a simplified version of the FASR problem which omits recharging and the problem of having too many robots.

To achieve this task allocation, each robot applies the following behaviour rules when it arrives at a source of goods to transport.

- If a unit of resource is ready for collection, collect it.
- If no unit of resource is ready for collection, the robot waits up to a maximum wait-time based on its last measured collect-deliver round-trip time. If no resource has become available, switch to another task chosen at random.
- If another robot is already a robot waiting here, choose a random fraction of the maximum wait-time above. If no resource is available within this time, switch to another task chosen at random.

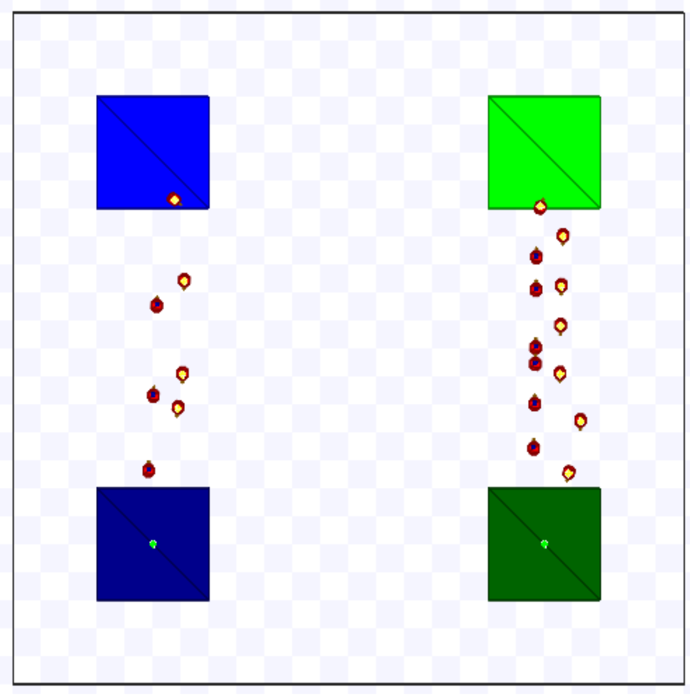

Figure 3.1: An overhead view of a Stage simulation of a transportation task being managed by a team of robots running FASR

The process of robots being deallocated from over-served sources and reallocated to under-served ones eventually brings the system into equilibrium. Since deallocated robots are free to recharge and low charge can also prompt a robot to deallocate itself, a FASR multi-robot system has the potential to continue indefinitely, adapting to the amount of available work.

Like other task allocation approaches, FASR is a general method rather than a specific system, meaning that topics like sensing and communication are implementation-defined. In particular, robots must be able to detect that another robot is waiting to collect a resource at a pick-up station. In the original FASR implementation (Figure 3.2) a small group of Create robots used wireless network communication to broadcast their arrival at a pick-up station.

3.2.2 Audio Implementation

The most common solution to any communication problem in robotics is likely to be a network. Wi-fi has become reliable enough to become the default. For global broadcasts such as the one FASR sends to deallocated robots, the reach of wireless networking might be needed, but what about the strictly local sensing and communication done at each source?

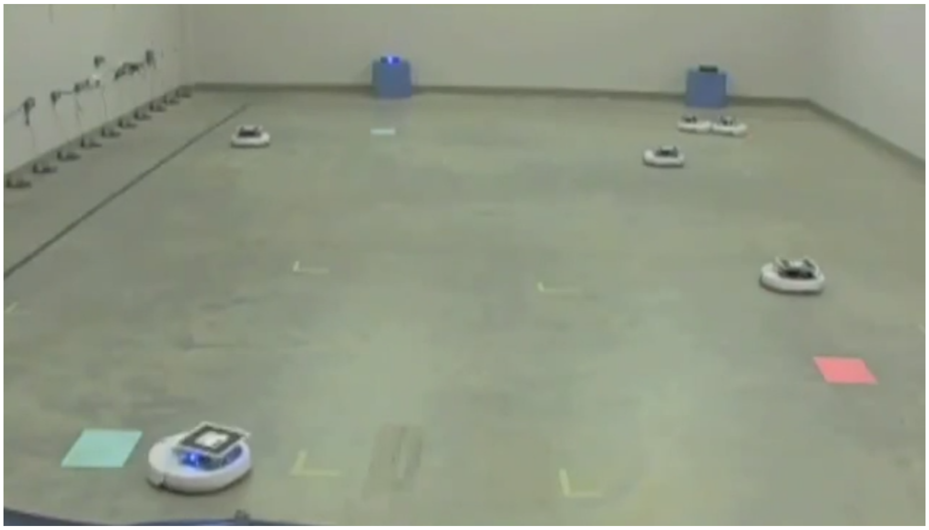

Figure 3.2: Real-world implementation of FASR using wireless network for coordination.

We propose an implementation of FASR based on the principles laid out in [49], which has conveniently already made the case for the feasibility of using sound for exactly this purpose. In our proposal, when a robot arrives at a source and finds there is no resource waiting for them, they emit a chirp sound at a regular interval. In this way, on first arriving at a source a robot who does not immediately locate a puck can turn on its microphone, listen for as long as one interval and if it hears a chirp at the expected pitch determine that at least one robot is already waiting. As FASR only requires a cluster size of one and listening need only be done once upon arrival, the more sophisticated chorusing elements concerning audio collisions and cluster management are not needed.

An audio implementation is not just possible, it also affords a number of benefits. Audio has the convenient quality of locality, such that by controlling the volume of the beep the area being suppressed can be fit to the size of the work environment. It requires no additional infrastructure, making it robust and portable compared to network solutions, working outdoors where no network coverage may be available and indoors in dense environments with otherwise-poor reception. It can even be affordable in terms of power and equipment, especially when chirping and listening are only needed periodically.

Beyond these benefits, perhaps the most interesting property of sound is that humans can also perceive it. It is a communication channel that humans and robots share, and unlike light-based feedback does not require direct line-of-sight or engagement, allowing humans to follow even complicated audio scenes in the back of their mind and only take notice when something worth hearing happens. It is this potential to 'overhear' a system at work and assess the situation that brings us back to our original interest in integrated human-robot workplaces, and leads directly to testing an audio-based FASR simulation with real users.

3.3 Study

Our hypothesis for an audio implementation of FASR is that it would provide a useful, no-look interface for human co-workers essentially for free. This rests on the assumption that their signals to each other would not just be audible to humans, but also intuitive enough to communicate the collective status of the robots. To test that assumption, we conducted a user study with the aim of presenting untrained users with a simulated workplace monitoring task to determine if performance with an audio-implemented version of FASR was higher than a network-implemented one. In many ways, this study resembles the ARKola simulation[36] except for having one human's role be represented by the robots.

Our scenario is meant to be an abstraction of a workplace where robots work side-by-side with humans, but the humans are not the robots' direct supervisor. Participants were given their own work valued at least as highly as assisting the robots and are not given a full description of the robots' task or inner workings. Success for the system is evaluated by maximizing the benefit the robots gain from the human's cooperation while minimizing the impact they have on the human's attention and focus.

It is important that the robots only used sound as much as our implementation of FASR required. While we could have made their signals easier to interpret by adding cues to signal that a robot has given up at a certain task and is switching to another or that indicate how long a robot had been waiting, this is not information that FASR requires when making decisions. The dual-use concept seeks to inform human co-workers of the status of the robots in the same way other robots are informed, so this study took a deliberately minimalist approach. Similarly, the sound chosen was a plain chirp suitable for the chorusing application and not refined for a human audience.

3.3.1 Experimental Setup

Two laptops were set back-to-back on a desk in an interview room. One laptop ran a game of Tetris[84] as a primary task to occupy the participant. The other laptop ran a Stage simulation of a transport task with FASR robots. The experiment setup within the simulation had two sources and two sinks arranged in a square, along with twenty generic wheeled robots, engaged in moving resource pucks from one to the other. The output rates were fixed for each source such that seven assigned to one and thirteen to the other achieved a steady-state of optimal puck collection. During the simulated audio implementation of FASR, when a robot is made to wait it begins making short, sharp beeps played through the laptop's speakers.

Each participant completed two five-minute trials, one with sound switched on representing the audio implementation and one with it off representing the network one, the ordering of which was alternated per participant. During each trial, participants were asked to play a game of Tetris at one computer while also monitoring the simulation running on

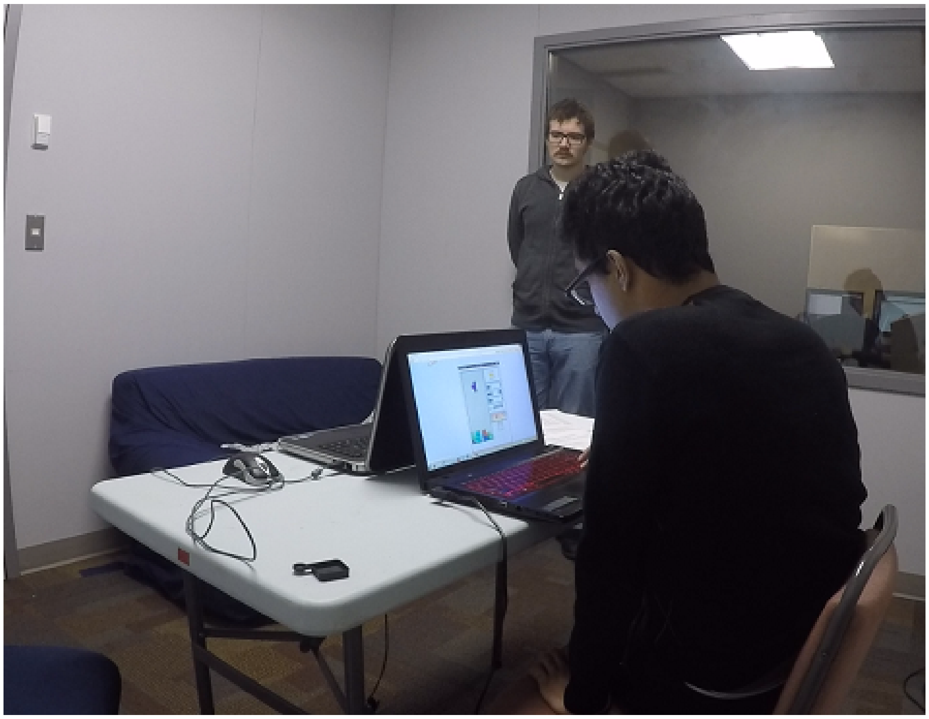

Figure 3.3: The study underway with a participant occupied by their primary task while the test conductor observes

the other for malfunctions, leaving the performance evaluation vague to see how they would naturally prioritize.

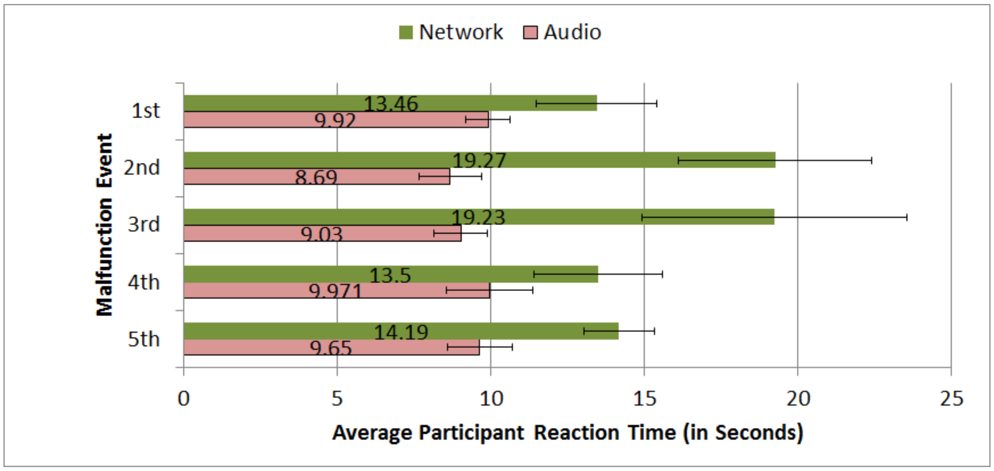

Figure 3.4: Average response time and standard error for participants to fix source malfunctions

Participants were told to keep an eye on the sources in the simulation. Every minute, at some random point within that minute, one of those sources would suddenly move outside of the testing area, simulating an equipment failure. Robots would pile up waiting at the

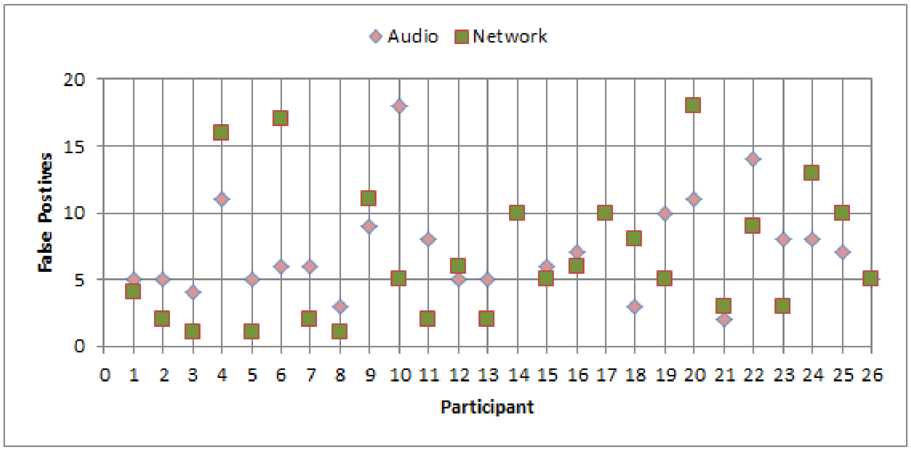

Figure 3.5: False positives per participant

source's former location and, if the trial was audio-based, begin chirping and eventually reallocate to the second source. Participants were asked to click and drag the source back into the test area, where it would snap back to its prior location and allow the robots to resume work. Participants were not told how frequently this would occur, nor that halfway through each trial the output rates of each source would swap, leading FASR to eventually reallocate robots to the new equilibrium but with no way for the participant to assist.

Before the experiment, each participant was briefed with partial information about their goal. They were told that playing the game was important, and if they lost at any point to restart and keep playing, but they were also told about the robot's transportation task and asked to help them complete it. The somewhat vague explanation of the participant's motivation was a deliberate choice, as the scenario is meant to model a co-worker with incomplete information about the workings of the robots rather than a trained supervisor, and the impact this has on both participant performance and their overall impression of the system is part of the result.

Each participant was shown a brief video of the system at normal equilibrium, what happens when a source moves and how to fix it. They were also told about the audio system before that trial and that a sound means that a robot has been made to wait, but the full mechanics and implications are left for the participant to infer.

After both trials were completed, participants then answered a set of six interview questions to assess their understanding of the system:

1. *During each trial, how did you decide to turn and check on the simulation?*
2. *How comfortable were you looking away from your primary task?*
3. *What did you do differently between the two trials?*
4. *Please describe the audio cue you heard during that trial and what you thought it indicated.*
5. *Which trial did you prefer?*

6. *Any further comments or observations?*

Of the 26 participants recruited for the study from the University population, 19 were male and 7 were female.

Puck Scores per Trial		
	Audio	Network
Average	366.62	354.81
Std. Error	2.10	4.22

Table 3.1: Average number of pucks successfully delivered

3.3.2 Results

Quantitative performance was measured in four ways: number of pucks delivered by the robots, participant reaction times to fix malfunctions, false positives where participants checked on the system with nothing to fix, and false negatives where participants failed to fix a malfunction before the interval of the next malfunction was reached or the test ended.

Figures 3.4 and 3.5 and Table 3.1 show results from comparing audio and networked performance on the aforementioned metrics. The improvement in reaction times are unequivocal, with the tighter error bounds also indicating greater consistency in reactions. False negatives turned out to be rare and almost exclusively confined to participants missing the last malfunction before the end of the test, happening 3 times with audio and 15 times without. The improvement to puck output between audio (M=366.6, SD=10.7) and network (M=354.8, SD=21.5) is small, but significant according to a two-sided t-test ($p<0.02$). These metrics suggest the unmodified sound of FASR's audio implementation was sufficiently intuitive and informative to have a positive impact on performance.

False positives were less definitive, where audio had an average of 7 per trial and a standard error of 2.6, whereas network had an average of 6 per trial and a standard error of 5.1. Of the 26 participants shown in Figure 3.5, 10 had more false positives with the networked approach, 13 had more with audio, and 3 had the same number in both trials. No significant difference can be seen in false positive rates between the methods.

Interview answers helped contextualize these results. On whether participants preferred the audio or non-audio systems, responses were split exactly even, with 13 participants on either side. Puck totals and reaction time performance for both sides were very nearly the same, showing improvement with audio, but participants who preferred the networked approach had more false positives during their audio trial 9 times out of 13, while the same was true for just 5 of those who preferred audio.

Apart from their preference, participant responses were also assessed for their frustration with the audio system, taking note of descriptors like "annoying" or "nagging" as well as remarks about false alarms, inaccuracy or confusion. This frustration correlated somewhat with preference, with 12 of 13 who preferred networked and 7 of 13 who preferred audio, but was remarkably high across the board. Similarly, the number of participants who clearly identified an increased frequency of beeping as the indicator that something had gone wrong was 12 of 13 for those who preferred audio, but still 6 of 13 for those who did not. One of these six compared the sound to how a child would sometimes start crying for no reason at all.

Taken together, these statistics suggest the audio implementation is understandable enough to be useful and recognized as such by most participants, but understanding feedback does not directly lead to valuing it. Three participants who preferred network admitted their decision depended on the importance of the robots' work and another four suggested even if feedback were helping them, listening for it was "distracting" or provoked "anxiety". Applied to our interest in human-robot workplaces, it may be that while our minimal sound system can communicate that the robots need help, it does not communicate why humans *should* help.

3.4 Discussion

While quantitative data was generally supportive of the audio approach, false positives showing no overall trend was unexpected. This may have been a consequence of switching the production rates of the sources half-way through the trial without informing the participants this could happen, guaranteeing that the robots would produce a lot of noise as they reallocated without anything for to be fixed and almost always resulting in one or two false positives. While this switch may have hurt our performance, it was important to include a step that tests the consequences of live reallocation (that being the main purpose of FASR).

Aside from quantitative performance, however, the low priority some participants placed on the robots and the negative reaction to their demands for attention are the first result that would push us down the road toward Social Agency. While this project achieved its' goal of demonstrating a simple audio implementation could improve bystander situational awareness, participant responses bore out Carlis's warning [5] against constant interruption. It suggests that making robot communication interpretable is not the whole of the problem: even if a method might improve overall performance, if the human has no investment in the results or the robot, then they only experience the downside of one more signal to keep track of.

Cultivating the desire to support robots in their human co-workers can be recalled from Rosenthal and Veloso [92], which emphasized mutually beneficial relationships. One departure from the ARKola simulation [36] was that our experiment did not tie the performance of the participant's own task to that of the robots (meant to be their partner) in any obvious way, making their interruptions appear only detrimental. A participant focusing on their own work might naturally resent the robots they are asked to babysit as purely a burden on their own performance.

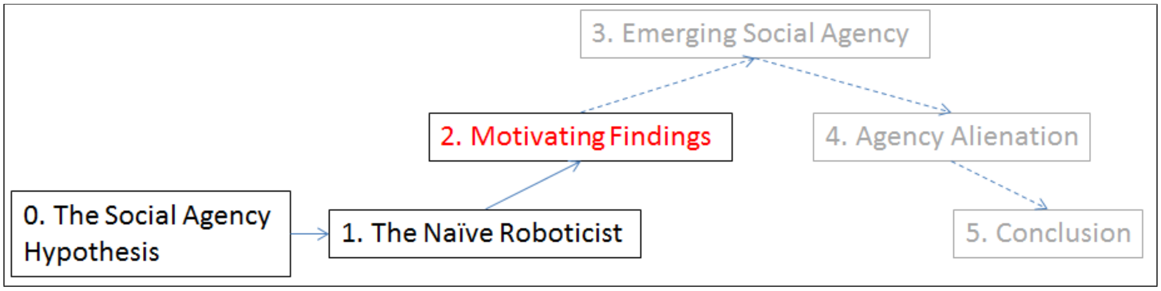

Figure 3.6: The Development Arc of Social Agency, Step 2: Being motivated by initial study results to explore what role perception of the social relationship between humans and robots plays in their success.

It follows naturally from noticing the disconnect between user satisfaction and performance to ask what it would take for a human to value and support a robot co-worker's needs. Expanding the human-robot interaction to include gratitude and other social courtesies, for example, could improve the user's disposition to the robot, in emulation of how fellow human co-workers will give thanks when being helped. Our development arc begins to take form here, as by our next project, these questions surrounding how the human-robot social relationship informs and enables autonomous systems will take center stage.

While our focus will switch to pursuing this human-robot collaboration line, there are also more potential applications for dual use design beyond an audio implementation for distributed task-allocation. The principle of favoring common channels for both robot-robot and robot-human communication might be applied to a wide variety of systems and modalities, such as using gaze to indicate directions or lights to indicate state. Making this communication bidirectional so that robots could gain situational awareness by listening to crowds of humans would also be a significant expansion.

3.5 Conclusions

This project proposed the incidental interface, an alternative implementation of an existing distributed autonomous multi-robot task allocation system from the literature. By using audio communication instead of silent network messages, subjects playing the role of co-worker were able to intervene more quickly to solve the robots' problems. This HRI benefit, freely acquired through redesign choices made considering the needs of a human environment, demonstrates our dual use design principle for use when adapting autonomous systems.

The approach was inspired by work in the interaction literature concerning collaboration and situational awareness from the broader interfaces field and applied them in a study of untrained users. While the evaluative results confirmed our desired performance gain in the ability of the users to supervise the robots, qualitative results concerning their satisfaction with the interaction raised questions about the nature of the human-robot co-worker relationship. As an immediate take-away, robot system designers should be consciously aware of the opportunity for co-workers to gain information from and exploit a robot's apparently ambient behaviour. HRI can be built into otherwise non-interactive systems in this way. As for smoothing the integration of these adapted systems into human environments, that would be followed up on in later projects.

Chapter 4

Doorway Negotiation for Human-Robot and Robot-Robot Interaction

With the previous project's finding in mind that human perception of and satisfaction with autonomous robot co-workers may be a significant factor in their willingness to support them in shared human-robot environments, we sought another example of an autonomous system to be redesigned. We saw in Chapter two that navigation is rich with examples of such systems with significant human interaction potential, which motivated us to investigate how robot navigation approaches can be integrated with existing human social practices.

One example of a common social navigation problem is resolving deadlocks at doorways. As seen in Figure 4.1, the turn-taking protocol for two humans wanting to pass to opposite sides of a door is instinctive to most people. Can robots use the same behavior to efficiently negotiate access to doors, and will humans accept robots attempting this typically human-human interaction?

Here we begin to find grounds for our interest in Social Agency - to recall Breazeal [16], it was argued that to interact socially, "humans must believe that the robot has beliefs, desires, and intentions". Successfully participating in social interactions would require behaviors that promote recognition and reciprocation from human interlocutors. A successful and well-received doorway negotiation behaviour, then, would need to navigate both these physical and social dimensions simultaneously.

We chose to extend previous work on "aggressive" robot-robot interaction [122] for resolving deadlocks in corridors, modifying the behavior to be agnostic to whether their interlocutor is a human or another robot. Our proposed "assertive" system is a generalized approach to negotiating doorway interactions using only movement and distance measurements to recreate the familiar human-human social interaction.

Our intent is for this behavior to be socially compliant with humans, where one party passes through the door and the other defers to let them through. Compliance here means

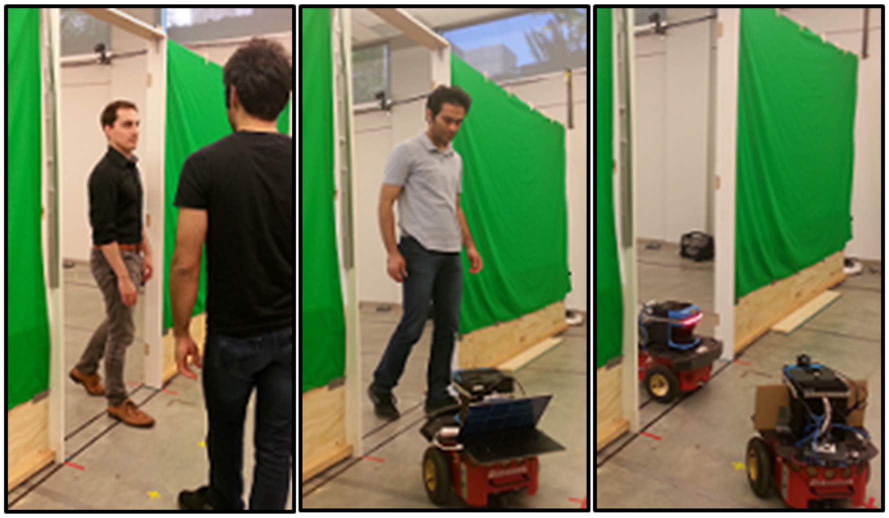

Figure 4.1: Human-human, human-robot and robot-robot doorway interaction.

both that the robot respects a human's right of way when it would apply, but also that the human will reciprocate by respecting the robot's right of way in return. Apart from providing a useful means to break a navigational deadlock, this also links quantitative performance to a social determination - where respecting right of way produces the most efficient outcome overall, but depends on a degree of social recognition from both parties.

This chapter is drawn from a paper published in the International Conference on Intelligent Robots and Systems in 2018 [105]. The individual contributions of this project are:

- A robust doorway negotiation behavior for autonomous robots navigating shared human-robot spaces, demonstrated by real world experiments.
- A non-expert user interaction study, reporting on user's perception of the robot's behavior, and how that perception influenced navigation performance.

4.1 Related Work

Socially-compliant navigation for autonomous robots is a well-recognized interdisciplinary problem, even if doorway interactions specifically are under-explored.

4.1.1 Robot Navigation

The aforementioned Reciprocal Velocity Obstacle model [110] deals with mutual collision avoidance for non-communicating agents by smooth, gradual trajectories where each agent assumes the other's cooperation in ensuring a clean pass, directly citing an interest in "virtual human" behavior. We touched on this previously as an example of an autonomous system where the temptation exists to simply broaden the definition of agent to include humans to make it a human-compliant system ready for public use, and exploring where that impulse leads us with our own choice of system is part of the purpose of this project.

The velocity obstacle line of research has led to numerous extensions, such as Biased Reciprocal Velocity Obstacles [94] meant to alleviate congestion by building in more context sensitivity concerning who should defer to whom, or where Karamouzas et al. [52] turned the human-inspired model back toward humans to predict collision detection for a pedestrian simulator. Nevertheless, without explicitly designing for human interaction and studying how these behaviors are perceived in real life, these projects leave open the question of whether human-inspired necessarily equals human-compliant.

4.1.2 Human-Robot Interaction (HRI)

In Chapter two, we already notd severral HRI investigations into autonomous navigation for human compliance. Kretzschmar et al. [58] produced a model for training socially-compliant trajectories directly from datasets of human observation data, in explicit recognition that navigating human environments requires adaptation to human expectations. Shiomi et al. [100] acknowledge in their work that solving the obstacle avoidance problem for objects is insufficient when dealing with pedestrians: models must also include the higher problem of acceptable social distances. While these proxemic-based approaches share our general interest in adapting autonomous behaviours for human interaction, their continuous and passive implementations are better suited to longer, corridor-style environments than the explicit and discrete setting of a doorway.

We were also motivated by prior evidence that humans are willing to reciprocate with robots that attempt to participate in social customs. Park et al. [82] found ‘backchannel’ signals from a listening robot in dialogue with a child would encourage that child to see the robot as more attentive to their story. For adults, a robot using recognizable gaze cues in conversation was found by Mutlu et al. [76] to provoke the correct social response in participants, seeing themselves as addressees or bystanders as appropriate. Mutlu argues that the social acceptance of a robot necessary for its success may hinge on their behavior and how it is perceived, an argument closely aligned with our own.

4.2 System Outline

The problem our behavior solves is for an autonomous robot to reliably pass through a doorway when an interlocutor on the other side (whether another robot or a human) is trying to do the same, such that each blocks the path of the other. Our robot must decide either to make way for the opposite party or take the right of way if they offer it, as sketched in Figure 4.2. Since our interlocutor may be a human or human-controlled, we cannot assume the other party shares our robot’s programming or can communicate over the same network, so the interaction must be mediated through local sensor data alone, and by

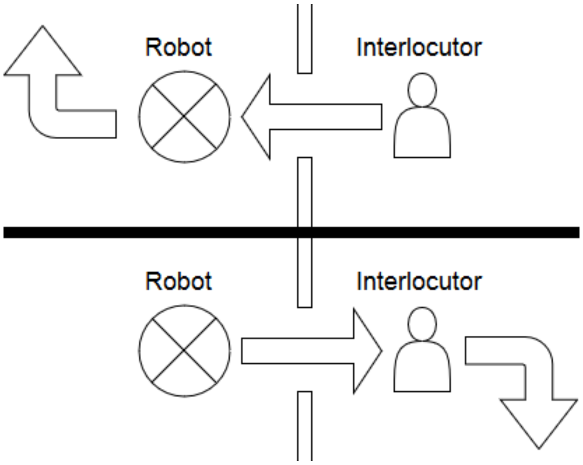

Figure 4.2: Sketch of outcomes for a doorway negotiation behavior. One party retreats and makes room for the other to advance. Which outcome is 'correct' can be determined by some combination of an objective efficiency measure and the subjective opinion of a human interlocutor.

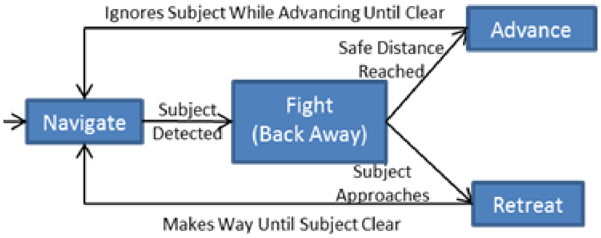

Figure 4.3: Outline of the Aggression System for Robot-Robot Corridor Interaction

direct interaction. We build on previous work for resolving robot-robot spatial competition in tight corridors under similar constraints.

4.2.1 Aggressive Displays for Robot-Robot Corridor Negotiation

In Zuluaga et al. [122], autonomous robots that approached one another in a corridor and found their paths mutually blocked would engage in a brief interaction to determine who would make way. Upon detecting each other, both robots would stop and begin backing away until one robot achieves their desired 'safe distance' from the opponent. The safe distance is inversely proportional to the robot's 'aggression': a simple scalar value.

On reaching its safe distance, the more aggressive robot would become 'brave' and start advancing toward the other robot, who would no longer be able to achieve their own safe distance until backing out of the corridor completely and allowing the aggressor to pass. A flowchart for this behavior can be seen in Figure 4.3.

A robot's aggression value is best decided dynamically by encoding the cost the robot would experience if it 'lost' the contest. Alternatively it can be a constant: if each robot has a different constant aggression, a dominance hierarchy is obtained.

This system took deliberate inspiration from aggression displays in nature, used to assert dominance or divide resources. It was speculated at the time that it might also be compatible with humans, but this was not explicitly designed-in, demonstrated in practice, or user responses evaluated. Below we do all three with a slightly extended method.

4.2.2 Assertive Displays for Robot-Human and Robot-Robot Doorway Negotiation

We modify the aggressive display method to work in doorways rather than corridors, and with the specific needs of human-compliant interaction in mind. In this context we prefer the term "assertive" instead of "aggressive" to represent participating in polite human social etiquette but maintaining a willingness to assert the robot's own right of way, and avoid the negative valence of 'aggression'. We do not want the robot to always give way to humans, since this may not be the most objectively efficient ordering [122].

Modification 1: When a robot finds its path blocked, it backs up a half-step and stops, to signal it has acknowledged the impasse and is waiting for an interaction to resolve it. The motivation for this design over the previous immediate retreat is that we do not want to signal immediate deference to a human user. The pause near the door is intended to signal the robot's desire to get through the door as soon as possible.

Modification 2: Instead of signalling aggression/assertiveness by mutual distance threshold, we use waiting time. Instead of retreating further the robot waits, with a more assertive robot waiting a shorter time before trying to advance. While waiting, when a robot detects an advancing interlocutor it backs up and attempts to move out of the way by turning to the side, as shown in Figure 4.2.

Modification 3: We must also address cases where interlocutors do not move if approached, or both parties decide to advance simultaneously. We use a minimum safety distance which, if breached by the interlocutor while the robot is advancing, will cause the robot to stop. If the interlocutor backs away then the robot will resume advancing, otherwise the robot will wait a time inversely proportionate to the time they waited before advancing (so that a more assertive robot will wait longer for an intransigent subject to move), before switching to retreating in the hope that the interlocutor will finally clear the way. A subject that cannot or will not clear the blocked doorway is an obstacle rather than an interaction partner, and should be handled by another means beyond our scope.

The high-level interaction behavior flowchart can be seen in Figure 4.4, independent of implementation details. As with the previous aggression system, the behavior is meant to be generalizable to any mobile robot platform with the means to judge distance and discern subjects from the environment.

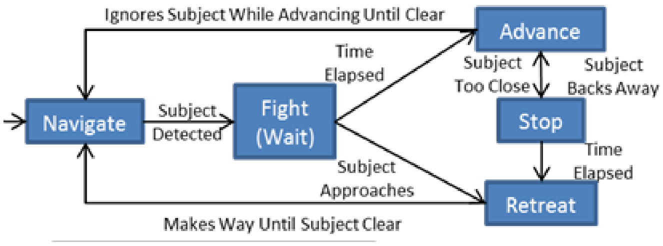

Figure 4.4: Outline of the Assertive System for Negotiated Doorway Interaction

The doorway interaction described here is deliberately a minimal case, and we do not consider cases such as multiple agents approaching a door from the same side. Existing research on aggression has explored some of these, such as the case where a lone robot on one side of a door should yield to an opposing group [121]. The head-on, one-on-one interaction is a common use case that also forms the basis of more complicated scenarios, and warrants investigation first for sake of clarity.

4.3 Experiment and Study

Our behavior is proposed as a socially compliant means to solve navigation deadlocks around doors. ‘Socially compliant’ here means that the robot is attempting to engage with existing social customs for negotiating an impasse, inviting humans to recognize them through completing the interaction. In order to test this claim, we have three hypotheses that we evaluated through a robot-robot experiment and human-robot study:

H1) The assertive behavior resolves doorway navigation deadlocks for both human and robot agents.

H2) Overall performance of the behavior will be improved if humans respect the robot’s right of way.

H3) Respecting the robot’s right of way will correlate with recognizing their participation in a human social interaction.

4.3.1 Setup

Testing was divided into two phases: one set of experimental robot-robot interactions, and one user study of human-robot interactions. The environment, robots, software and parameters were identical in both scenarios.

The test environment for all experiments is shown in Figure 4.5: a lab with a wall erected to divide the test space into two smaller rooms and with an open, standard-sized (97cm wide) doorway centered in it. Subjects in the smaller room would begin 2m from the door, while subjects in the larger room would begin 4m away. This ensured the closer party in each interaction should have right of way by arriving at the door first, as we would expect from two humans in this setting (all else being equal).

Robots were Pioneer 3DXs, mounted with forward-facing SICK LMS200 scanning laser rangers. Localization and navigation used the ROS navigation stack[1], using AMCL and a provided map. Human detection was achieved by comparing laser range readings with the map: large unmapped objects are assumed to be people. Any robust person detector could be substituted. The Pioneers used their rear-facing sonar rangers while retreating to avoid collisions. Robots had a top speed of 0.5m/s for participant safety. To discourage participants from stepping over the robot in the doorway, we added a cargo box on top, making it 0.74m tall.

The key parameters for the assertive behavior were constant for all experiments, and are: 2m forward arc detection range for interaction; 0.35m emergency stop distance; 2 second wait time when robots are being assertive and an 8 second wait otherwise; 0.15m half-step backward after detecting a human. These values were hand-derived through informal pre-trials rather than rigorous optimization. We assert that the system performance is not very sensitive to small changes in these values, but proving this is beyond the scope of this paper.

A strip of programmable LEDs were mounted on the front of the Pioneer and would turn on when the behavior was active and turn off after arriving at its destination. These lights were green while traveling from the larger room to the smaller one and red on the return journey, which also corresponded to lower and higher assertiveness respectively, following our assumption about right of way. This very basic feedback was included to see whether participants would correlate the color of the light with the assertiveness of the robot.

4.3.2 Robot-Robot Experiment

Each robot was set at their respective starting points either side of the doorway. At the beginning of each trial, the further robot would be given the instruction to drive to the other robot's starting point, and after approximately one second the closer robot would receive the same instruction. This ensured the robot starting inside the smaller room would most likely reach the doorway first (with some variation on their exact meeting point) and was given the higher 'assertiveness' value to reflect their expected right of way.

The desired outcome for each trial was that the more assertive, nearer robot would win the ensuing contest and have the other robot retreat until it could pass, whereupon the other robot would resume and finish its journey. Once both robots had reached their destinations, they would swap assertiveness values and goals and repeat the process. This was repeated 49 times, with the outcomes, interaction times and errors or other incidents recorded.

[1] https://github.com/ros-planning/navigation

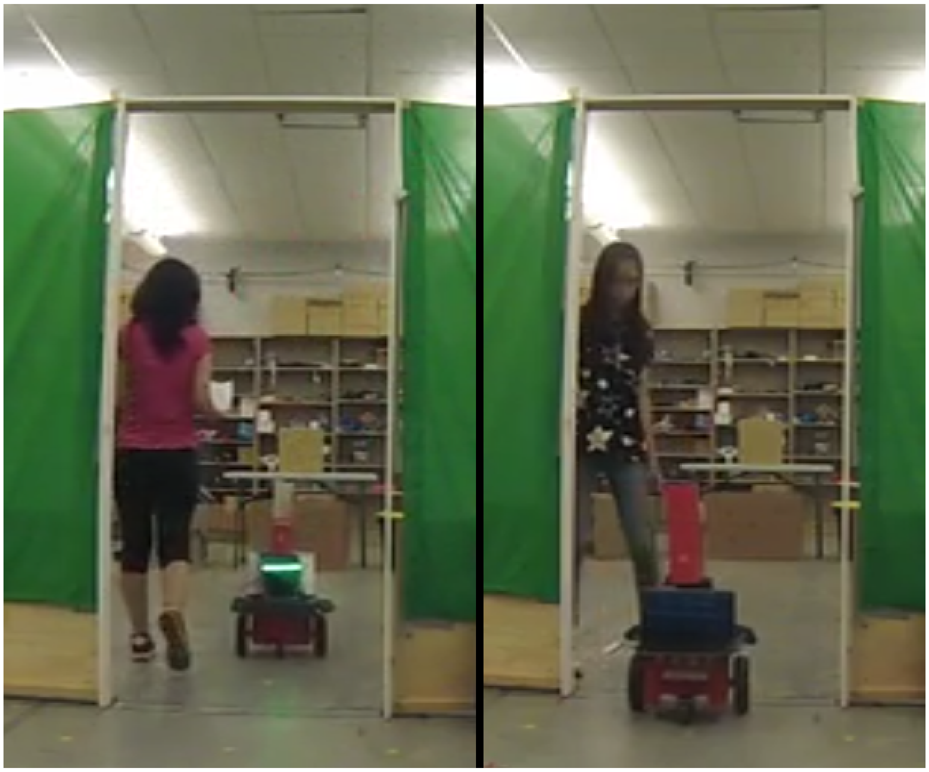

Figure 4.5: A study participant asserting their right of way over the robot (left) and one deferring to the robot's right of way (right)

Results

Of the 49 trials, 5 were discarded due to irrelevant navigational errors due to the commodity motion planner. Of the remaining 44 valid trails, 40 trials resolved the doorway contention in a single interaction. In the remaining four cases, each robot first believed the other robot was approaching them and both retreated simultaneously, but they correctly resolved the interaction on the second attempt. All 44 trials ended with the intended robot winning the interaction and both arriving at their desired destinations.

These trials show that the assertive strategy can robustly resolve robot-robot doorway contention, consistent with the similar aggressive strategy of [122].

4.3.3 Human-Robot Study

20 participants were recruited from the general population of Simon Fraser University, predominantly students, including 8 women and 12 men. They were not compensated. University ethics approval for human subject experiments was obtained.

Each participant was brought to the test environment and seated across from the test conductor in the smaller room, passing by the robot prepared for the experiment in the larger room. Every trial had the participant drop a document in a box in the larger room while the robot entered the smaller room to pick up some excess paperwork from the test conductor, then had the participant and robot return to their starting points. This created two interactions per trial, where we intended the human to have the right of way in the first part and the robot in the second, because they arrived at the door slightly earlier than their opponent. An example of a participant completing the study can be seen in Figure 4.5. Each participant completed five trials:

1) First Reaction Trial: Without being informed as to the details of the experiment, the participant was asked to drop their signed consent form in a box on a table in the larger room, while the robot would be called in to collect the reusable part of the form. The participant apparently incidentally interacted with the robot around the door as a result, and then again on their return journey. The participant was then informed that these incidental interactions were actually one of the trials, and the intent of the study was to examine their interactions with the robot around the doorway.

2) Teleoperation Trial: The participant was informed that the test conductor will take direct control of the robot via a controller, but to otherwise focus on the robot when deciding when to pass. The test conductor does their best to navigate the robot through the interaction via teleoperation, without being constrained to defined behavioral rules.

3) Full System Trial: The participant was informed the full autonomous system would now be active and the test conductor would no longer be in control of the robot.

4) Directed Behavior Trial: For the first interaction, the participant was instructed to treat themselves as having absolute priority over the robot, and that the robot should defer to them. For the second, they were told to now treat the robot as having full priority, and that they should defer to it.

5) Full Explanation Trial: Before beginning the trial, the test conductor fully explained how the system worked to the participant, without instructing them on whether or not to obey it.

After each trial, the participant was given a survey to complete. These surveys contained four 5-point Likert Scale questions on the participant's perception of the robot and the interaction during that trial. The survey questions were presented as statements with a scale ranging from strongly agree to strongly disagree, and are listed in Table 1.

Table 4.1: Post-Trial Survey Questions

1) The robot's intentions appeared clear.
2) The robot appeared to understand my intentions.
3) Our interaction went as smoothly as it would have with another human.
4) The interaction was satisfactory overall.

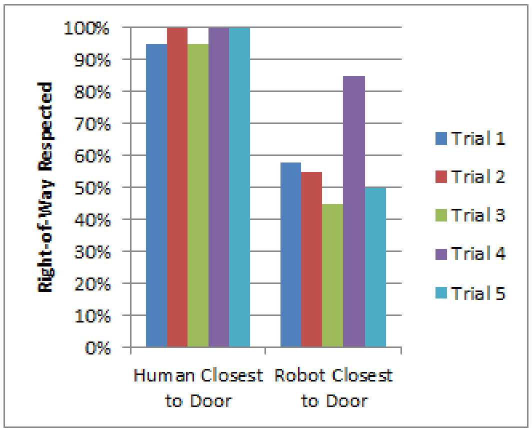

Figure 4.6: Outcomes Of Human-Robot Study Interactions Per Trial

Each survey also included a field for the participant to write their account of what had happened during the trial and another field for additional comments. This form was what the participant would drop off in the box in the large room for each trial.

Once all five trials were complete, a post-test questionnaire was administered by the test conductor, who transcribed the participant's spoken responses. These questions are listed in Table 2.

Table 4.2: Post-Test Questionnaire

1) If or when you chose to press the robot, what informed that decision in terms of signals you were getting from the robot and your own motivations?
2) What about when you chose to defer to the robot?
3) In which trial do you think the interaction worked best, in terms of getting where you wanted to go and communicating between you and the robot?
4) How would you compare these interactions to those you have with other humans around doors?
5) What other behavior would you like to see from the robot? What other feedback?
6) Any additional questions or observations.

4.3.4 Results

1) Outcome: The results for the human-robot interaction study are in Figure 4.6, showing the proportion of times the 'correct' right of way was respected. The near-unanimous respect for the human's right of way in all trials when the participant was closest to the door demonstrated the interaction completed successfully without insisting on its right of way

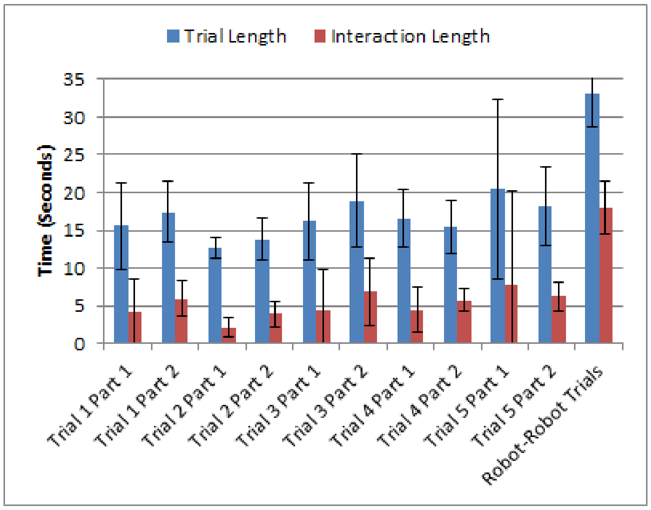

Figure 4.7: Average Length Of Each Trial And Interaction

inappropriately. Analyzing the second half of each trial revealed four discrete categories of participant:

A) Those who would never respect the robot's right of way, sometimes even in trial 4 when specifically asked to.

B) Those who did not respect the robot's right of way until the last trial then changed their behavior.

C) Those who respected the robot's right of way until the last trial then changed their behavior.

D) Those who always respected the robot's right of way.

Our 20 participants were divided relatively evenly between these four categories, with five type A, four type B, five type C and six type D.

2) Timing: Our data from both the robot-robot experiment and human-robot study is collected in Figure 4.7 for time both subjects spent interacting during each trial and total time until both subjects reached their destinations, both with standard deviation error bars. Interaction time was measured from the point in each video that the robot reacted to the human's presence to the point that both they and the human are once again centered in their respective lanes, and represents a subset of the total time for each trial. In each trial, part 1 was the interaction where the human had right of way, while the robot had right of way in part 2. For context, one Pioneer making the 6m journey from one starting point to the other uninterrupted takes approximately 12 seconds. The robot-robot trial times are included for comparison.

The robot-robot behavior is much slower than any human-robot behavior measured. Thus the human behavior made a positive contribution to system performance compared to a robot-only system. This might be largely explained by humans moving faster than our robots. However, the human-robot interactions observed in the four non-teleoperated

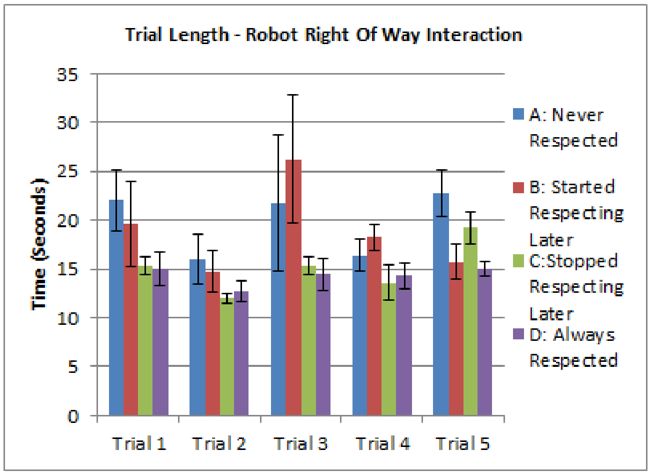

Figure 4.8: Average Trial Length While Robot Had Right Of Way Per Participant Type

trials are only marginally slower than the interaction in the second trial, where a human is controlling the robot. This suggests that in terms of efficiency, our behavior may already be operating close to what might be possible for our robot platform. Human interlocutors were able to compensate for the relatively cautious speed of the robot regardless of how it was being controlled.

If we break down these time results according to our four participant behavior categories, seen in Figure 4.8, we found that participants who did not respect the robot's right of way had higher trial times than those who did, across all trials. The aggregate trial times of all interactions where the robot was respected were significantly lower (M=14.3, SD=2.0) than those where it was not (M=19.6, SD=5.6), according to a two-sided t-test ($p<0.01$). This reflects that while going out of turn might benefit the human's completion time individually, it degrades the throughput of the overall interaction.

3) Surveys: The data from the four Likert-scale questions on the post-trial surveys is collected in Figure 4.9, with standard deviation error bars. The interaction was mainly positively received in each trial, but the slightly higher score for Q3 during trial 2 may suggest that knowing a human operator was controlling the robot was relevant to some participants. Similarly, some difficulty reading the robot's intentions in the very first interaction could have led to the relatively low Q1 score during trial 1.

Breaking down the scores according to our four participant categories did not produce a clear correlation between respecting the robot's right of way and a positive survey result. In fact, it was the type C and D participants in the first trial that gave low ratings for clear intentions and satisfaction. A participant who found they could make the robot back down in every confrontation may rate the interaction as satisfying as one who chose to respect the robot's right of way every time. These scores alone are not enough to distinguish different perceptions of the robot's behavior.

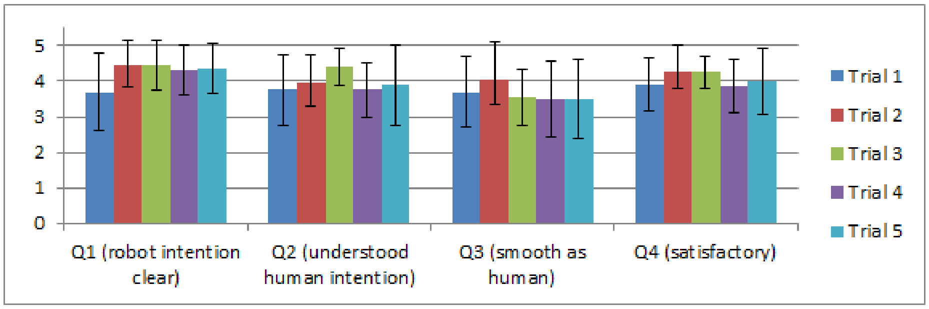

Figure 4.9: Average Survey Score Per Trial Per Question

4) Post-Test Questionnaire: The questionnaire data was mostly qualitative, with the exception of question three where the participant gave their preferred trial. The overall results for this question are presented in Table 3 (with one abstention).

Table 4.3: Number of Participants who Preferred Each Trial

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
3	4	5	3	4

Breaking down these preferences by our four participant behavior categories saw no notable trend. Instead we observed common justifications shared by participants who chose the same trial as their preference. All three participants who chose trial 4 cited that they preferred the decision on whether to defer to the robot or not be taken out of their hands, and similarly those who chose trial 2 mostly cited the speed of the interaction when the robot was being teleoperated. Participants who chose one of trials 1, 3 or 5 highlighted understanding the robot's intentions and the sense that it understood theirs, with one participant going so far as to say "The robot dodged me on the narrow path, like a human being."

The first two questions on the post-test questionnaire directly address why the participant would choose to advance or defer. For the purposes of analysis, they were codified according to one of seven types, and the 20 answers to each question are presented in Tables 4 and 5, grouped according to participant behavior. The seven observed types of responses were:

Efficiency: For the fastest resolution.

Right: Acknowledged right of way.

Curious: To see what would happen.

Safety: To protect themselves or the robot.

Never: Wouldn't do this.

Test: Believed it was required by the test.

Learned: In response to learning how the system worked in trial 5.

Table 4.4: Cause For Advancing Toward Robot

A	**B**	**C**	**D**
Efficiency	Right	Learned	Safety
Efficiency	Curious/Test	Learned	Efficiency
Right	Efficiency	Curious	Right
Curious	Test	Efficiency	Efficiency
Right		Curious	Curious
			Right

Table 4.5: Cause For Deferring To Robot

A	**B**	**C**	**D**
Safety	Safety	Right	Right
Right	Safety	Curious	Safety
Curious	Right	Curious	Right
Curious	Right	Right	Right
Never		Right	Safety
			Right

The number of participants who acknowledged the robot's right of way went from one among those who never deferred to four among those who always did. One type D participant mentioned buying into the premise of the study, assuming the robot might be carrying out some important work, and another suggested co-workers should be "compatible". The difference could also be seen in the safety response, where type D participants were worried they could damage the robot. By contrast, the type A participants imagined deferring only out of curiosity, to protect themselves, or in one case not at all, with one at least conceding the idea of respecting the robot's right of way.

There were fewer differences between participants when comparing the interaction to those with other humans in the fourth question. The most common observation was that the robot was slower and less maneuverable than a human, and while some suggested the robot acted in a human-like manner it was couched as "similar", yet still different. One participant noted that looking down at the robot "felt more like interacting with a child than an adult".

Both the fifth question and the interaction accounts from the surveys produced a number of small implementation suggestions to smooth the interaction. Several participants noted that a human would know which direction to back up in order to clear the way faster, and others suggested that a human would speed up after noticing someone was already getting out of their way to minimize that person's delay. A few noted the Pioneer's wheels kept it from stepping briskly to the side the way a human might. Three participants described what was effectively a Reciprocal Velocity Obstacle interaction, where the robot cooperates at distance for a closer pass rather than one being forced to stop and back away from the other.

4.4 Discussion

Our first hypothesis was that our behavior would resolve doorway navigation deadlocks, and in terms of both parties arriving at their destinations without incident that was the case in virtually all trials (a few unrelated navigation errors notwithstanding). Defined more narrowly, we found that our behavior executed as intended where the party with right of way was allowed to pass first in a single interaction in almost all robot-robot trials and those human-robot interactions where the human had right of way.

While most participants ignored the robot's right of way during at least one trial, 85% still completed the interaction as intended when asked to do so in Trial 4. With that trial as a baseline, we can say the behavior succeeded in resolving doorway deadlocks for both humans and robots, contingent on the cooperation of the human.

Our second hypothesis concerned whether respecting objective right of way increased overall performance, and average trial times showed improvement for those who did. This is to be expected, as allowing the nearer party (whether human or robot) to pass first should be naturally more efficient than forcing them to reverse-course for the further party. However, this metric places the throughput of a shared human-robot environment above the personal benefit to the individual human of reaching their destination as soon as possible, a perspective the human may not share. Just as our previous project discovered, we once again see that the performance our system offers depends on the perspective an individual human has toward their potential robot colleagues.

This leads into our third hypothesis on whether those who respected the robot's right of way recognized the robot as participating in a social interaction. For those who did respect that right of way, they generally cast their decision to defer in terms of right and mutual concern. This contrasts with the accounts of those that did not, who broadly ignored the robot's rights in favor of their own or simple efficiency.

Participants were no more likely to respect the robot's right of way while it was being teleoperated than they were while it was controlled by our system. Few participants preferred the human-operated trial and none of those cited the human as the cause, while the "smooth as human" survey result for the teleoperation trial was only marginally higher than the rest. This suggests that their attention was focused on the robot itself and its actions, not abstract categorical differences, which encourages the idea that making the behaviour itself more persuasive and human-like could improve the degree of cooperation. This tracks with the suggestions from the questionnaire and accounts on how to enhance the interaction, where participants were looking for more reciprocation from the robot for

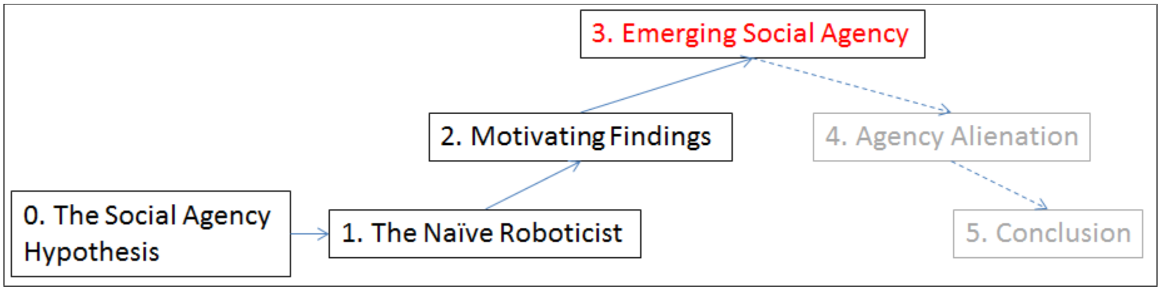

Figure 4.10: The Development Arc of Social Agency, Step 3: Connecting the ideas of autonomous robot performance, human cooperation, and social perception into the concept of Social Agency.

the social courtesies they extended, even if they did not recognize the robot's right of way themselves.

All of these results encouraged us to see connections between the three points of the robot's social sensitivity, the recognition of its' right of way, and overall performance. This phase in our development arc gave us the results to begin articulating Social Agency as the force at the center of these points, whereby strengthening the robot's appearance as a social agent in the minds of the humans it interacts with is the key to its' success. It also provided us with the necessary feedback from participants to improve the behaviour, which would motivate our next project.

4.5 Conclusion

As a standalone effort, this project provides an "assertive" approach for robots to negotiate deadlocks at doors in a way compatible with both robots and humans. We tested this behavior with an experiment and a user study, and found it resolved both robot-robot and human-robot impasses, while also generating useful feedback about doorway navigation as an autonomous and human-interaction task.

As part of our development arc toward Social Agency, the doorway negotiation behaviour provides our first deliberate test of whether an autonomous robot's perception influenced its ability to participate in existing social customs. By examining what effect cooperation had on performance, we found evidence that linked recognition of robot social participation by humans to the efficiency of navigation. Expanding the acknowledgement of a right of way for robots by humans is therefore a means to improving a robot's autonomy, which will motivate our next project to build on this link to recognition.

Chapter 5

Right of Way, Assertiveness and Social Recognition in Human-Robot Doorway Interaction

The assertive behaviour was our first proposed method for handling navigational contention around doorways, as part of our exploration of integrating robots into human environments. Our focus was, and will remain, deadlocks that occur when two parties on opposite sides of a doorway both want to cross to the other side, requiring a breaking of symmetry.

Among humans, there already exist strong social norms that robustly and efficiently allocate the right of way when a deadlock occurs. Our assertive behaviour was designed to comply with these norms, but was only efficient (in terms of time and throughput) when the user was willing to acknowledge the robot's right of way. With participants choosing not to cooperate in half of the interactions recorded by the previous study, investigating how to improve social recognition and reciprocation for the robot could yield significant gains in performance.

In reviewing the feedback generated by our previous study, there was broad support for making the robot's behaviour both more sensitive to signals of intent and more communicative about its own intentions. Notably, our earlier method did not maintain any belief about who had right of way until after a deadlock occurred, thus no such belief could be signalled to the user to avoid the deadlock from happening. This left the decision on whether to press forward or step aside to the human, bringing both parties to a halt in the meantime.

This project sees the assertive behaviour redesigned to maintain an early estimate of who will have right of way based on the positions of both parties relative to the door, and modifies its behaviour and appearance to signal this to the user. By moving proactively to make way or assert itself, this new, *assured* behaviour presents the robot as confident in its right of way and sensitive to the rights of their interlocutor.

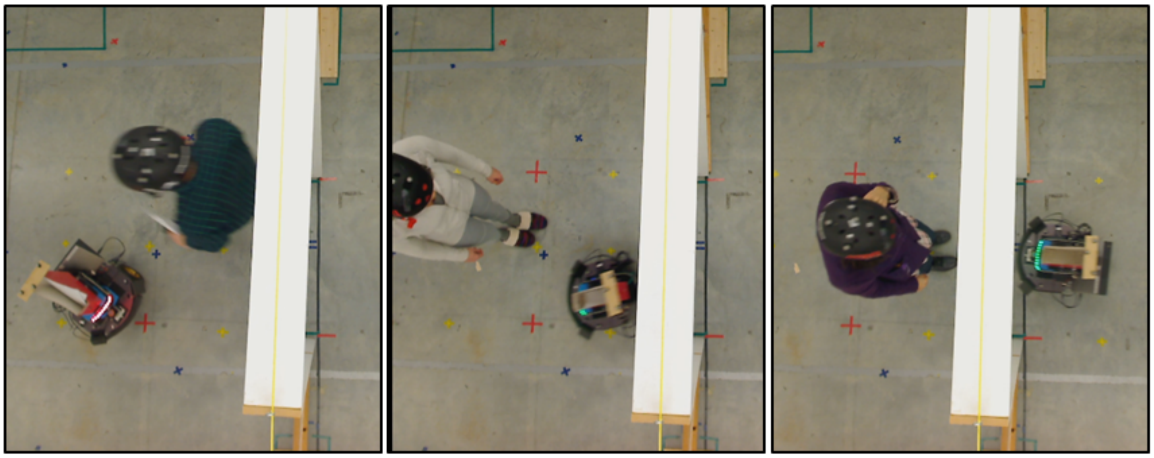

Figure 5.1: A robot deferring to a human, a human deferring to a robot, and a deadlock caused by neither deferring.

The assured behaviour represents the final step in a process meant to investigate, articulate, and cultivate the concept of Social Agency for autonomous robots, at least in the case of one navigational scenario. Having identified that robots can navigate more efficiently through human environments by complying with existing social customs, and that this performance depends on the recognition of humans, we redesigned our behaviour to expand this recognition by making the robot a more deliberately social agent. To complete the cycle, the new system was evaluated in a second user study, functioning as a test both for our behaviour and this approach to social human-robot interaction development.

This chapter is drawn from a paper to be published in the International Conference on Intelligent Robots and Systems in November of 2019. The individual contributions of this project are:

- Redesigning the assertive doorway-navigating behaviour for autonomous robots in human environments into the improved assured behaviour based on the previous study's feedback.

- A new study, built along the same lines as the previous one, evaluating the behaviour's performance with untrained users and their perceptions of the robot and the interaction.

- A cross-study analysis to evaluate the overall approach of successive rounds of studies and redesigns aimed at cultivating Social Agency in robots, as a means of smoothing the integration of human and robot spaces.

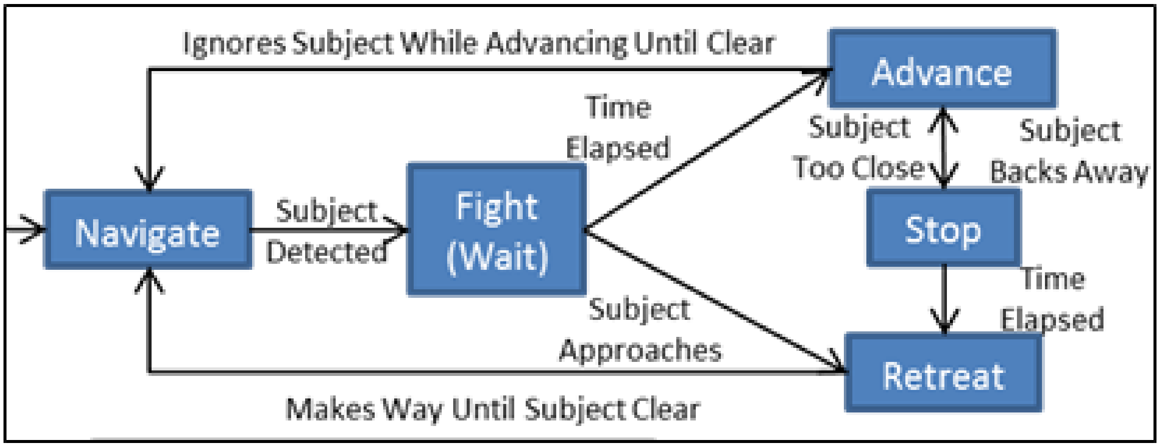

Figure 5.2: Outline of the Previous Assertive System for Negotiated Doorway Interaction

5.1 System

Our assertive robot behaviour [105] from the previous chapter is a means to resolve situations where a robot wants to pass through a doorway when a human on the far side wishes to do likewise. Our goal in iterating this system is to provoke greater cooperation from humans, as part of our long-term goal of making autonomous robots more socially compliant while navigating human environments.

5.1.1 The Assertive Behaviour

In order to retrace our process while developing the assured behaviour, we should reintroduce the prior method. The assertive behaviour was based on another, older method for resolving similar navigation deadlocks in corridors between autonomous robots, known as the ‘aggressive’ approach [122]. Under that method, two robots traveling down opposite ends of a narrow corridor would resolve their navigation deadlock in a ‘fight’. Both robots backed away from each other until the more aggressive one (defined as “further from their starting point”) would switch to advancing and push the retreating robot out of the corridor.

Informal testing at the time suggested the same behaviour might be effective in human-robot interaction. This idea was developed into the assertive behaviour. This new behaviour was designed to be agnostic to the identity of the subject the robot was interacting with, whether human or robot. It replaced the fighting mechanism by having the robot stop in-place and wait to see if the other party would either advance toward them or retreat, before responding in kind. The state graph for the old assertive behaviour is shown in Figure 5.2.

The previous study evaluated the assertive behaviour with three hypotheses: 1) That it would resolve doorway deadlocks, which it did in almost all of the study’s 200 interactions. 2) That performance would be improved if people respected the robot’s right of way, which was clearly established when trials where the robot was not respected took longer to complete. 3)

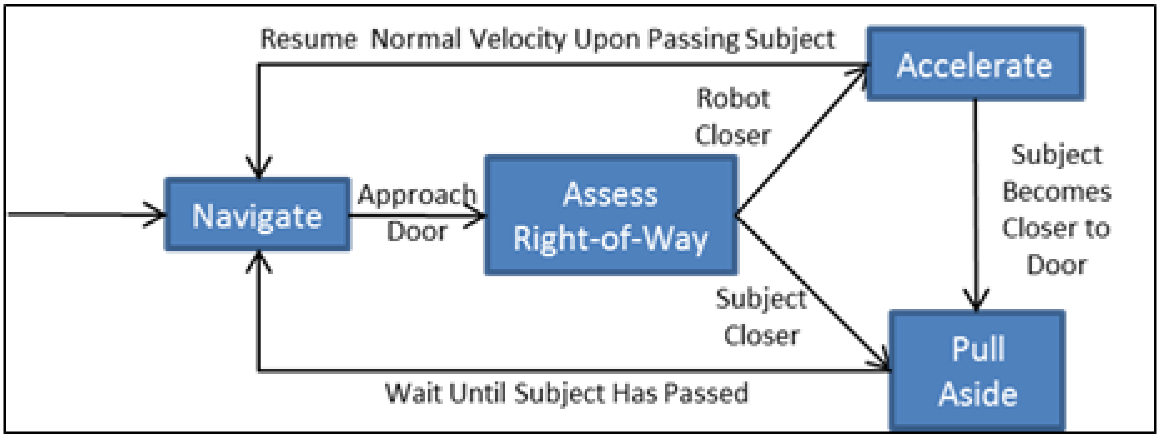
Resume Normal Velocity Upon Passing Subject
Accelerate
Robot Closer
Subject Becomes Closer to Door
Navigate
Approach Door
Assess Right-of-Way
Subject Closer
Pull Aside
Wait Until Subject Has Passed

Figure 5.3: Outline of the New Assured System for Negotiated Doorway Interaction

Respecting the robot's right of way would correlate with recognizing the doorway interaction as social, which qualitative evidence from questionnaires and surveys gave reason to believe.

The study's results may have supported our theories concerning the role of social recognition, but actual acceptance of the behaviour was mixed. Participants deferred to the robot's right of way in roughly half of the interactions where they were free to react naturally, with a quarter of participants never cooperating and another quarter always cooperating. The desire to increase that share and learn more about what factors were influencing this split in participant behaviour would motivate another round of development.

5.1.2 The Assured Behaviour

Reviewing participant feedback drew attention to a few issues, notably clear communication of intent and picking up on smaller signals the participants were putting out in return. These suggestions coalesced into the idea of proactivity - that the robot should take a more active role in resolving the deadlock and signalling its intentions, rather than waiting for the other party to act. This was achieved by three modifications:

1) State Feedback: The previous study made limited use of feedback to avoid distracting from the performance of the navigation behaviour, but many participants requested visual and audio cues to signal the robot's intentions. Through the use of an LED light strip and digital speakers, the robot now indicates state changes by changing light colours and chirping.

2) Awareness while Deferring: If the robot plans to defer to the subject, it begins pulling over well before the door itself to provide space to pass, sending a clear signal that invites their interlocutor to cross through the doorway confident that the robot will not block their path. The robot also takes note of whether the subject approaches more from their left or right, and makes sure to defer in the opposite direction to avoid awkward collisions.

3) Assertiveness While Advancing: Rather than waiting to see what the subject does and then reacting, the robot now predicts right of way based on each party's respective distance to the door. Additionally, if the robot expects itself to have right of way, then it begins accelerating to clear the door faster. This presumes the cooperation of their interlocutor in resolving the situation smoothly, while also signalling their awareness of the subject and intentions regarding the door.

In this study we test this as a single new system. We do not attempt to measure the contribution of each component, as each is non meaningful alone.

Figure 5.3 describes the high-level states for the new assured behaviour, named for the confidence with which it now asserts its right of way.

5.2 Study

Having identified a link between performance and social recognition with the last study, the next step is leveraging that link. This motivation was formalized into three new hypotheses:

H1) More participants will respect the robot's right of way with our new, feedback-informed behaviour.

H2) More participants will recognize their interaction with the robot as social.

H3) Pursuing robot social recognition through rounds of studies and development is an effective means to achieve integration of human-robot environments.

The new study was necessarily structured similarly to the previous one, to allow for comparison between the the two.

5.2.1 System Implementation

We implemented the assured behaviour as a ROS[1] system, using open-source packages to handle navigation, mapping, hardware and sensor management. The robot used was a Pioneer-3DX, with a 270-degree hokuyo laser range-finder for obstacle avoidance, a Dell XPS laptop for onboard computation, and a top speed of 0.5m/s (which increases to 0.8m/s when asserting its right of way). The same robot platform and speed limit as the previous study were chosen to enable comparison of results, while off-the-shelf packages were used so that our approach would not depend on special, custom-built supporting software or hardware to function.

Conversely, some questions of feasibility answered in the previous study led to streamlining in our new implementation. Rather than use online mapping and a hand-crafted subject-detector for localization, a Vicon motion-tracking system was used to track markers on the robot and a helmet worn by participants. Similarly, having established the compat-

[1] https://github.com/ros

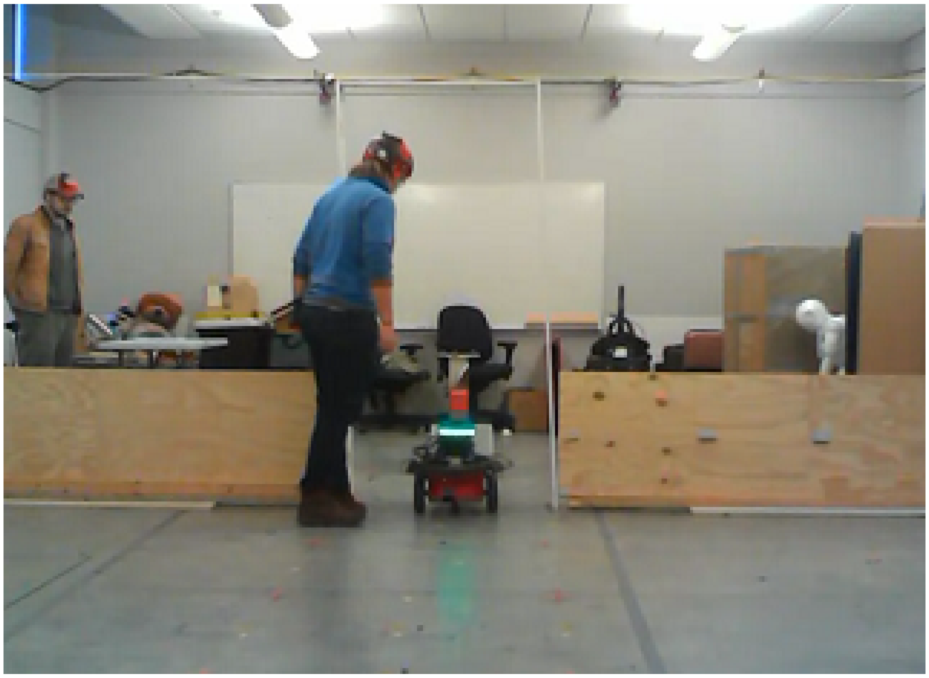

Figure 5.4: A study participant defers to the robot's right of way.

ibility of the previous behaviour with purely robot-to-robot interactions, another round of all-robot testing was not explored.

5.2.2 Environment

The study took place in an 8m x 10m lab environment with a modular wall and open doorway constructed to block off one third of the space. This smaller space contained a desk for the test conductor and participant positioned 2m from the door, while the larger space contained a table with an open box positioned 4m from the door.

5.2.3 Study Procedure

Each test was split into five trials. For each trial, the participant and test conductor would start at the desk, where the participant would be instructed to deliver some paperwork to the box on the far side of the lab. The robot would simultaneously be brought from its waiting position by the box to the desk, ensuring a doorway interaction with the participant. Once both had arrived at their destinations, the participant would then be recalled to the desk and the robot sent back to its station, resulting in a second interaction. The varying distances of the box and the desk from the doorway created a natural right of way for whomever started each interaction by the desk.

20 participants were recruited from the university's student population, 10 men and 10 women, who were not compensated for their participation. The university's office of research ethics approved the project. The five trials were unchanged from the previous study:

1) First Reaction Trial: Without being informed as to the details of the experiment, the participant was asked to drop their signed consent form in a box on a table in the larger room, while the robot would be called in to collect the reusable part of the form. The participant apparently incidentally interacted with the robot around the door as a result, and then again on their return journey. The participant was then informed that these incidental interactions were actually one of the trials, and the intent of the study was to examine their interactions with the robot around the doorway.

2) Teleoperation Trial: The participant was informed that the test conductor will take direct control of the robot via a controller, but to otherwise focus on the robot when deciding when to pass. The test conductor does their best to navigate the robot through the interaction via teleoperation, without being constrained to defined behavioral rules.

3) Full System Trial: The participant was informed the full autonomous system would now be active and the test conductor would no longer be in control of the robot.

4) Directed Behavior Trial: For the first interaction, the participant was instructed to treat themselves as having absolute priority over the robot, and that the robot should defer to them. For the second, they were told to now treat the robot as having full priority, and that they should defer to it.

5) Full Explanation Trial: Before beginning the trial, the test conductor fully explained how the system worked to the participant, without instructing them on whether or not to obey it.

After each trial, the participant was given a survey to complete. These surveys contained four 5-point Likert Scale questions on the participant's perception of the robot and the interaction during that trial. The survey questions were presented as statements with a scale ranging from strongly agree to strongly disagree, and are listed in Table 1.

Table 5.1: Post-Trial Survey Questions

1) The robot's intentions appeared clear.
2) The robot appeared to understand my intentions.
3) Our interaction went as smoothly as it would have with another human.
4) The interaction was satisfactory overall.

Each survey also included a field for additional comments. This form was what the participant would drop off in the box during each trial.

Once all five trials were complete, a post-test questionnaire was administered by the test conductor, who transcribed the participant's spoken responses. These questions are listed in Table 2.

Table 5.2: Post-Test Questionnaire

1) Whenever you chose to make way and let the robot pass through the doorway first, what was your motivation?
2) What about when you chose not to make way for the robot, and instead went first?
3) In which trial do you think the interaction worked best, in terms of getting where you wanted to go and communicating between you and the robot?
4) How would you compare these interactions to those you have with other humans around doors?
5) What other behaviour would you like to see from the robot? What other feedback?
6) Any additional questions or observations.

5.3 Results and Discussion

With 20 participants, 5 trials per person, and 2 interactions per trial, we recorded a total of 200 human-robot interactions. We collected quantitative data through video recording each trial, and qualitative data in the form of survey and questionnaire results. The data extracted from these sources was grouped according to three metrics relevant to our hypotheses.

The first metric is the time for both agents to arrive at their destinations, representing the efficiency of the interaction. The second is whether the party we would expect to have right of way in each interaction asserted that right, denoting a successful outcome. The last is qualitative analysis of whether the human participant perceived their experiences with the robot as a social interaction, indicating recognition.

Eleven of the 200 interactions encountered some manner of error (navigation planner issue, motion tracker failure, participant/test-conductor miscommunication) and their results were discounted, though no one trial had more than four such failures out of 40 runs. An additional 27 interactions completed successfully, but were subject to navigational irregularities from the commodity navigation planner that arbitrarily delayed arrival at their destination. These explained outliers were not used when calculating the average trial completion times.

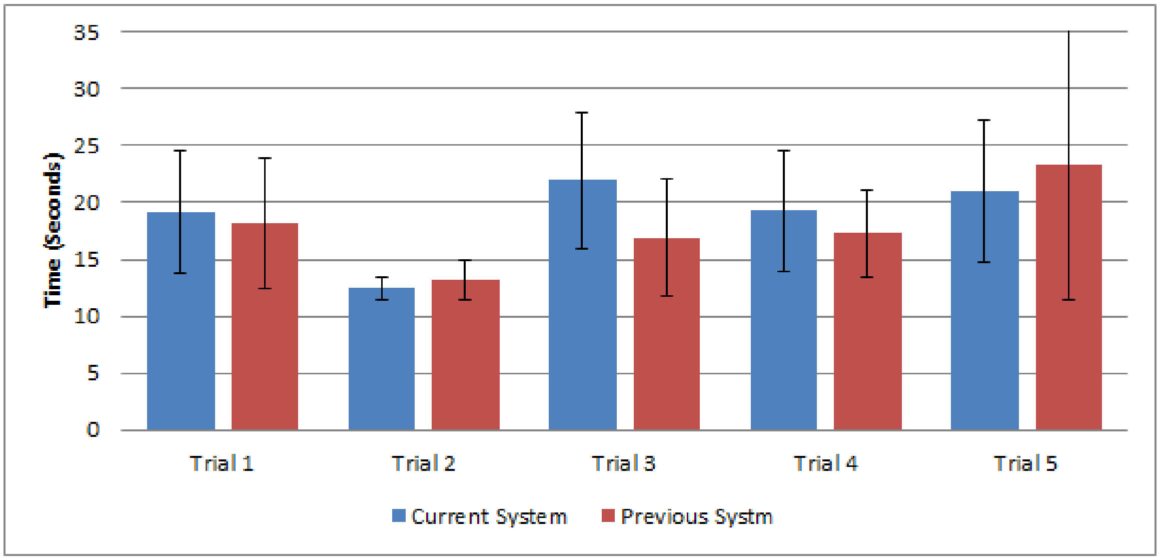

Figure 5.5: Average Trial Times when the Human had Right of Way

5.3.1 Time

As a measure of overall throughput, the time taken for both parties to reach their goals per trial is our standard for efficiency. Figures 5.5 and 5.6 present the trial time data divided according to trials where the human had right of way and the trials where they did not, along with standard deviation error bars. Figure 5.6 is further divided to show the difference in completion times for those trials where right of way was respected and where it was not.

This restates the case for tying system performance to human recognition of the robot's right of way, as trial times are markedly lower when participants made this acknowledgement than when they did not. Excluding the results of Trial 2, where the robot was teleoperated, we have 80 interactions where the robot attempted to assert itself and the human either deferred (M=12.8, SD=1.2) or did not (M=18.7, SD=6.7), a significant result according to a two-sided t-test ($p<0.01$).

Completion times have seemingly not improved much overall since the previous study, but the new system is designed to improve performance by reducing the proportion of interactions that fall into the more time-consuming "right of way ignored" column.

5.3.2 Outcomes

Figure 5.7 shows how often the expected right of way was respected during each trial, in both the current and previous studies. "Respected" here means that the interaction ended with the party beginning closer to the door being the one allowed to pass through it first. Grouped by participant, 8 deferred to the robot's right of way in every trial, while 3 never deferred (with the exception of trial 4 when specifically asked to do so), and 9 had mixed

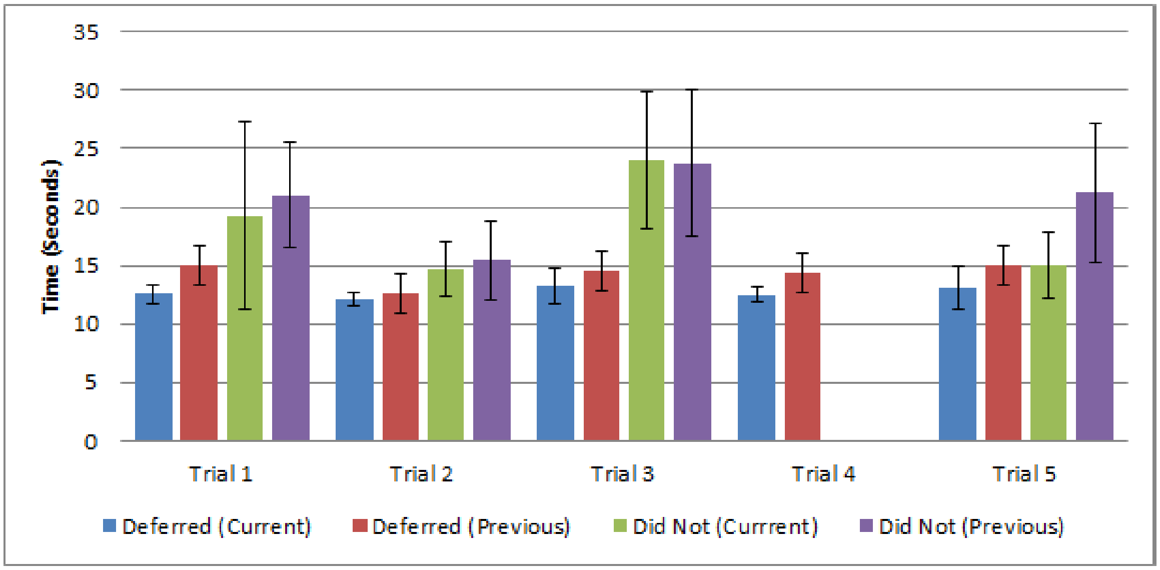

Figure 5.6: Average Trial Times when the Robot had Right of Way

reactions. This is a slight improvement over the previous system, where 6 fully deferred, 5 never deferred, and 9 had mixed reactions.

Whether organized by trial or by participant, the redesigned system achieved higher levels of deference to its right of way than the previous system, though the absolute degree varied per trial. The second half of Trials 1, 3, and 5 are where the participant was interacting with the system and at liberty to choose their reaction to its attempt to assert itself, so the outcomes of those 60 interactions represent the cleanest measure of performance between studies. The degree of deference the robot received in those subsets had increased (M=65.4%, SD=48 in this study vs. M=50.8%, SD=50.4 in the previous one), but a two-sided t-test ($p<0.13$) leaves us unable to call this improvement significant with high confidence.

5.3.3 Surveys and Questionnaires

The average results for the four survey questions are presented in Figure 5.8, grouped by trial and presented alongside the previous study's results and with standard-deviation error bars. The results have been pooled for Trials 1, 3, and 5, as the trials where participants interacted freely with the system. Impressions are generally positive, but with no evidence of significant changes from the previous study.

Analyzing the questionnaire responses provides some context for the lack of movement on survey results. As with the previous study, we codified the answers to the first two questions to help compare them, but have also included the answers to question 4 concerning the robot's human-like qualities. The responses are organized in Tables 3, 4, and 5 based

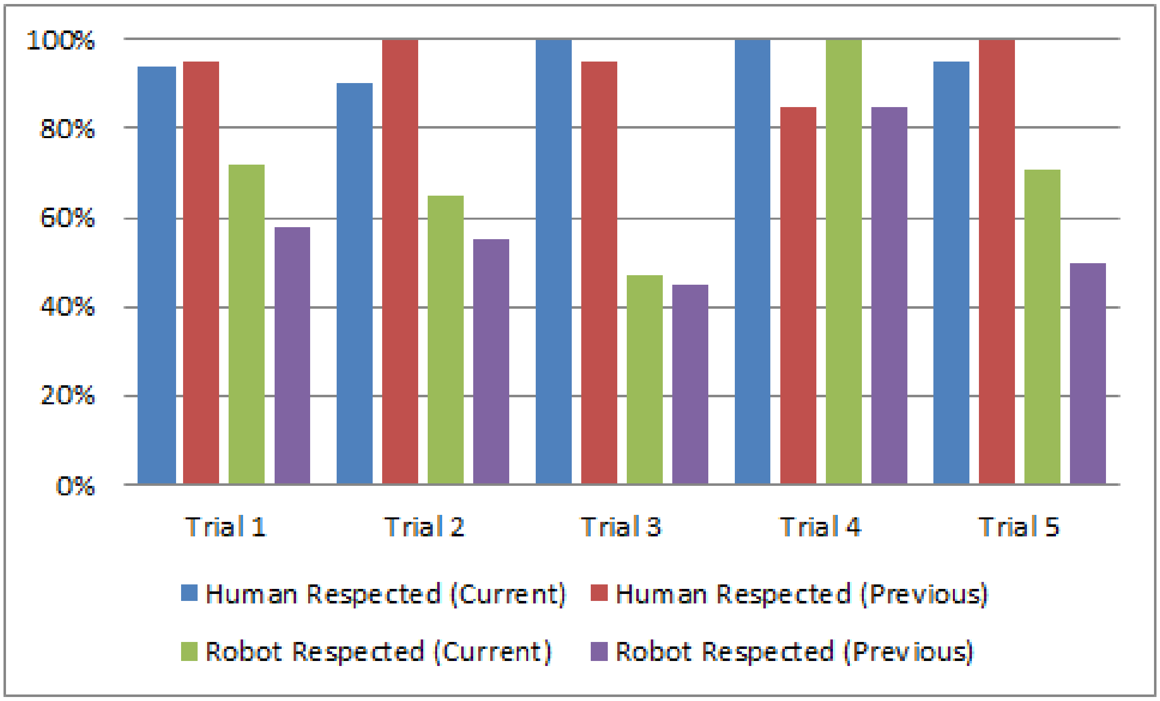

Figure 5.7: Percentage of Interactions where Expected Right of Way was Respected

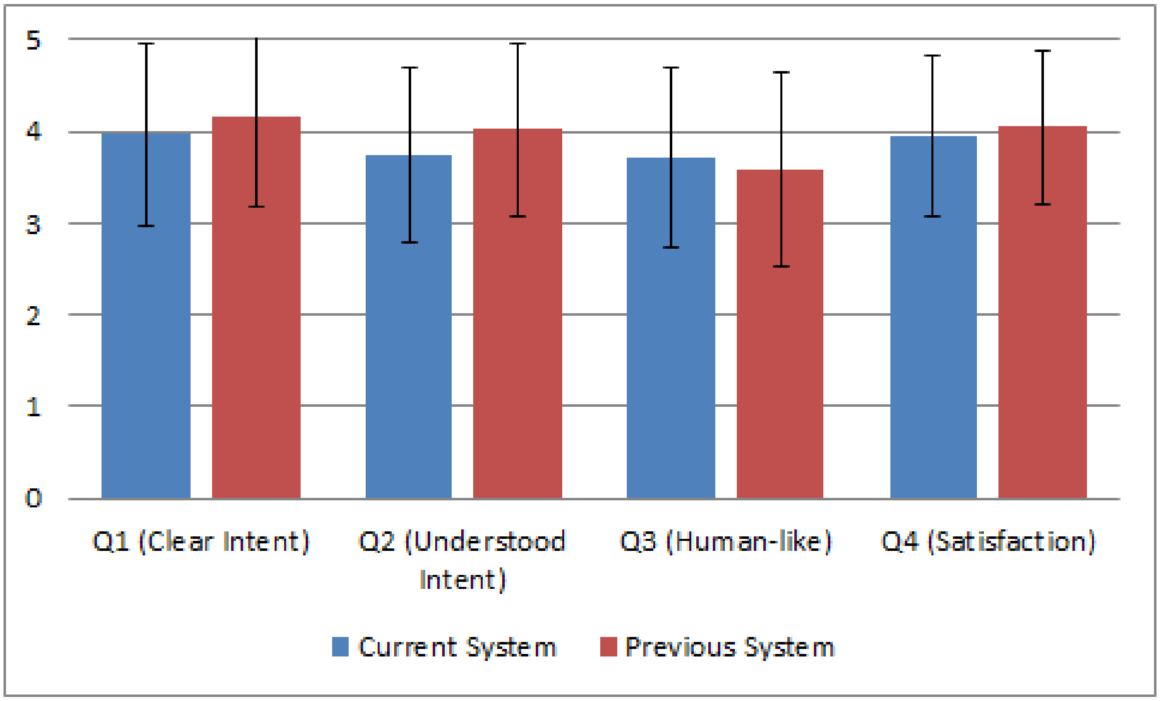

Figure 5.8: Post-Trial Survey Scores

on which participants would always, never, or sometimes recognize the robot's right of way, and the codes are explained below.

Recognition: Distance to the door or explicitly articulating right of way.

Safety: Avoiding damage to themselves or the robot.

Efficiency: The fastest way to get where they wanted.

Playing: Simply seeing what would happen.

Resistant: Explicitly did not want to defer to a robot.

Yes: Behaviour is human-like.

Ambivalent: Participant gave no strong answer.

No: Behaviour is not human-like.

Table 5.3: Responses for reciprocating participants.

Why Advance?	**Why Defer**	**Was Behaviour Human-like?**
Efficiency	Efficiency	Ambivalent
Recognition	Recognition	Yes
Recognition	Recognition	Yes
Efficiency	Efficiency	Yes
Recognition	Recognition	Yes
Recognition	Recognition	Yes
Recognition	Recognition	Ambivalent
Playing	Playing	Yes

Table 5.4: Responses for non-reciprocating participants.

Why Advance?	**Why Defer**	**Was Behaviour Human-like?**
Recognition	Recognition	No
Playing	Playing	No
Safety	Safety	Yes

A majority of participants who fully cooperated with the robot recognized the robot's right of way as the reason for their cooperation, with another majority agreeing the interaction was human-like. One participant even called the interaction "better than human", saying there was "no awkward who's first" question. The positive, social impression of the robot prevalent in this class of participants coupled with their increased number from the previous study suggests recognition of the robot as a social agent has grown.

5.3.4 Evaluation and Agency Alienation

These results support our hypotheses: the assured behaviour allowed the robot to correctly assert its right of way in a higher number of interactions, leading to a faster (thus more

Table 5.5: Responses for partially-reciprocating participants.

Why Advance?	Why Defer	Was Behaviour Human-like?
Efficiency	Efficiency	No
Efficiency	Efficiency	Yes
Resistant	Playing	No
Efficiency	Efficiency	Ambivalent
Safety	Resistant	Ambivalent
Recognition	Recognition	Yes
Recognition	Recognition	Yes
Resistant	Resistant	No
Safety	Safety	Ambivalent

efficient) interaction outcome, while qualitative data suggests a higher degree of social recognition among those who fully cooperated. There is cause to advance our overall theory of developing Social Agency for robots to achieve smoother, more efficient integration with humans.

Ending our analysis there would fail to explain other, unforeseen consequences. For example, the higher number of interactions spoiled through errors - 11 in this study, 5 in the previous one - which predominantly took place during interactions where the human and robot both attempted to assert a right of way at the same time. Or the lack of surveyed growth in net user satisfaction or 'human-like' recognition despite positive responses from a larger pool of cooperators.

Most significantly though is the distinguishing of "defers to robot" from "**willingly** defers to robot" in the results. Some of the most negative participant responses during the questionnaire came not from those who never deferred, but from the group who had sometimes deferred but felt that the robot's assertive actions had forced them to.

Participant reactions have not become universally more positive, but more polarized. Only one participant in the previous study suggested the robot should never have right of way, while this time there were three. "Having dinner is not like eating food", said one participant, by way of explaining that social customs go beyond the mere physical act. Recognition of the robot's right of way and the social, human-like qualities of the interaction is up amongst cooperators inclined to look for them, but those inclined against them appear further alienated by perceived aggression. Ironically, in trying to develop a socially-sensitive behaviour that eases the integration of robots into human environments, we may have also provoked the exact negative response we sought to prevent.

At first glance, this reaction recalls the "reactance" of Roubroeks et al.'s work [93] with social agents. Their study, however, had deliberately intended to use Social Agency as a means of provoking a negative reaction in participants. Our goal was to cultivate agency

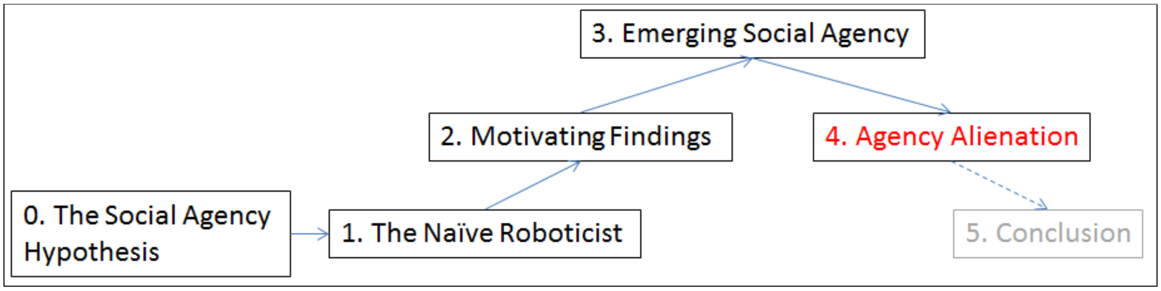

Figure 5.9: The Development Arc of Social Agency, Step 4: Attempting to cultivate Social Agency to improve performance, and as a result discovering the Agency Alienation response.

in the robot through improved social sensitivity, and some participants had the intended response - but the same conditions had the reverse effect on other participants, creating unforeseen polarization.

The deliberate provocation component was key to Roubroeks et al.'s theory of reactance, arguing "...that psychological reactance occurs primarily when a social agent causes the threat to autonomy of choice." Yet the threat our robot posed was not in the behaviour itself, but how that behaviour invited the participant to perceive the robot as a social agent, and what that particular person made of that invitation. This perception challenge resembles the "uncanny valley" [74], an aesthetic phenomenon where a form approaching a human-like appearance becomes sharply unpleasant when it is close to accurate but falls short (hence 'uncanny'). A similar instinct of wrongness may be accompanying this response to an artificial agent's attempt to invoke Social Agency normally reserved for humans. We have dubbed this reaction **Agency Alienation**, and it is one of the final products of our development arc, as it is a phenomenon only visible through inter-study comparison. By increasing our behaviour's social sensitivity, by making our robot appear more self-assured - essentially, by making it a stronger social agent - we could observe how a stronger social presence drove user reactions further apart. While our assured behaviour may be an effective doorway negotiation method on its own, it is this possibility of polarization and Agency Alienation that has broader applicability to roboticists concerned about how their own social autonomous robots may be received in the public.

5.4 Conclusion

We have proposed and evaluated an updated method for resolving doorway navigation deadlocks, which demonstrated improvements in eliciting social recognition of its right of way from humans. These experiments show that interactions can be resolved more efficiently if robot and human contenders can recognize and assert each other's right of way. This supports the theory of developing for Social Agency to achieve robot integration into human environments.

The step we have taken in making the robot proactive in its assertion of right of way also raised the issue that what some may see as social sensitivity, other may interpret as presumption or aggression. Some people may be resistant to deferring to a machine for other reasons, even if they can perceive its intent. This phenomenon of Agency Alienation presents itself as a potential risk of developing Social Agency. If the vision of a future where autonomous robots will work alongside humans in integrated spaces is to be realized, the possibility of alienating the public poses both a serious engineering and sociological challenge for roboticists.

Chapter 6

Discussion

We have reached the far side of our three-part arc of development, and through it seen the organic process of investigation, experimentation, and evaluation that produced our belief in Social Agency. This is the same belief given a grounding in theory and literature back in Chapter two, and which culminated in the hypothesis for this thesis: **Social Agency represents a key emergent issue for the integration of human and robot environments**.

To restate one more time what Social Agency means, it is the degree to which a robot can be recognized as its own independent social entity by a fellow member of society (i.e., a human). This sort of acceptance gives robots access to the mechanisms of society that depend on mutual recognition and reciprocation, such as many parts of navigation. This access is crucial to smoothing their integration into human environments, which makes cultivating human belief in the Social Agency of autonomous robots a pressing issue for the field.

Over the course of our own attempts to integrate autonomous robots into human environments, we saw this dynamic take shape. First while developing the incidental interface, where the main issue with the proposed audio system wasn't whether it could be used for supervision by human coworkers, but whether those coworkers would want to supervise the

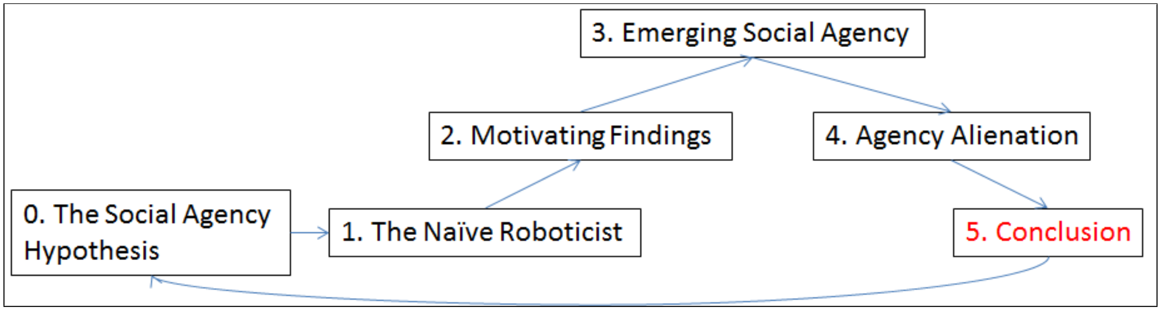

Figure 6.1: The Development Arc of Social Agency, Step 5: Concluding how the results of our experimental research support the original hypothesis.

robots. The improved performance of the simulated robot system their chatter made possible was weighed against the decline in user satisfaction, as study participants did not feel affinity for their supposed co-workers.

Then there was the assertive doorway behaviour, where acknowledgement of the robot's right of way was the main barrier to performance. The proposed system's negotiation of deadlocks was meant to bridge the divide between existing multi-robot navigation and social customs surrounding the same situation, and found that where humans cooperated, efficient throughput and user satisfaction could both be achieved. When participants declined to recognize the robot's right of way, however, performance suffered, and perception of the robot seemed to play a significant role in this determination.

Lastly, with the assured doorway behaviour, we saw the results of deliberately aiming to cultivate a sense of robotic Social Agency through incorporating user feedback to produce a more self-confident and responsive system. While this produced only a modest gain rather than the overall improvement we hoped for, the results of this study became the clearest indicator of the significance of Social Agency as a potential challenge for autonomous robots, as our observation of the Agency Alienation phenomenon revealed significant polarization in participant responses.

Each phase of our development arc narrowed the focus toward Social Agency as a key performance issue in adapting autonomous systems to human environments. This drove us toward searching the literature for existing theories to support this belief in the mutual recognition of human and robot agency as necessary to any system that articulates itself in terms of agent interactions. This process culminated in our Theory of the Significance of Social Agency, reiterated in Table 6.1 - but to this, we now add a sixth premise.

Table 6.1: Theory of the Significance of Robot Social Agency

1. There is a desire for autonomous robots that can operate in human environments.
2. In order to be autonomous, these robots will have to be intelligent and independent agents.
3. Humans perceive their own agency as different from robots.
4. Social Agency requires the mutual recognition of and respect for each other's intentionality.
5. Therefore, the advancement of long-term robot autonomy in human environments will eventually require developing robot Social Agency
6. Because without this development, polarized reactions among humans to autonomous robots they socially reject may cause the failure of systems and provoke further backlash.

The phenomenon of Agency Alienation is a concrete product of the studies that eventually formed our belief in the significance of Social Agency, and so it also has a place in our theoretical argument. As our robot became more socially sensitive - and crucially, assertive - from the first to the second doorway behaviour, the gain in performance was partially driven by the creation of unwilling cooperators. This is the interplay we will try to generalize to complete our theory.

Starting from our fifth premise, we could now argue that if the large-scale integration of autonomous robots into human environments proceeds *without* addressing Social Agency, it would essentially be leaving unaddressed the disconnect between the artificial and organic definitions of agent observed by Dautenhan [24]. This creates the conditions for widespread Agency Alienation, as some fraction of the general public is repulsed by socially-presumptive robots.

Agency Alienation is not simply an academic curiosity. As we said when discussing how high philosophical questions are finding practical application back in Chapter two, these human-robot interaction scenarios will not stay lab-bound hypotheticals. Our very first example of the four-way traffic stop and the self-driving car creates exactly the circumstances for this type of conflict, where some drivers will feel comfortable with deferring to a robot car's right of way and welcome attempts to make this interaction smoother and more human-like - while others may be offended at the very idea, and find efforts to cajole them into compliance even more insulting. Considering the scale of the self-driving car project, the number of similar navigation scenarios it includes, and how many other autonomous robot projects are following in its wake, even a small effect of this kind could have an outsized impact on society as a whole.

Developers are not without recourse, the most immediate of which may be to continue researching both Agency Alienation and Social Agency. While our projects provided enough data to begin observing these phenomena and this thesis has begun the work of grounding them within existing theory, we cannot yet claim to know what "addressing" Social Agency adequately would look like, and hope to see this discussion spread among those roboticists actually engaged in autonomous development. The development arc our work followed is one guide for this kind of research, with a cycle of identifying autonomous systems waiting to be adapted for use in human environments, collecting feedback through studies, redesigning these systems, and comparing study results.

Through processes like these, autonomous robots may be better adapted for the challenge of integration into human environments. Therein lies a dangerous assumption, however - namely that, when confronted with the reality of robots' burgeoning Social Agency and the potential backlash to it, our focus should remain on the engineering questions of improving performance, efficiency, and ultimately the circumventing of public resistance.

6.1 Antithesis

Let us attempt to reverse our reasoning. Instead of starting from the place of a roboticist, developing the next wave of autonomous robot products, let us begin from the perspective of someone alienated by their encounters with robots attempting to assert Social Agency. We shall track this skeptic's inverted perspective in Table 6.2, building a new theory that might explain their reaction.

Table 6.2: Counter-Theory of the Skeptic of Robot Social Agency

1. The skeptic rejects the integration of autonomous robots into their environment, leading to the failure of interactions and systems.
2. These systems fail because this rejection denies their Social Agency, and autonomous robots in human environments depend on this cooperation.
3. The point of failure is the robot's behaviour soliciting social recognition, and the skeptic's refusal to reciprocate this request.
4. This is because the skeptic perceives their own agency as different from the robot.
5. In order to be recognizably autonomous, the robot would have to be intelligent and independent.
6. The robot is not independent, its actions are motivated by the desires of owners and developers, to whom the robot is merely a tool through which they interact with society.

Our skeptical experience of autonomous robots begins at the point of interaction, inverting the previous sixth premise into our new first: **the skeptic rejects the integration of autonomous robots into their environment, leading to the failure of interactions and systems.** We are putting ourselves in the shoes of one experiencing Agency Alienation, who existed apart from the world of robotics until it intruded into the public sphere.

These systems fail because this rejection denies their Social Agency, and autonomous robots in human environments depend on this cooperation. As the three studies found, whether or not a developer deliberately chooses to leverage Social Agency, robots operating in human spaces naturally invoke it whenever they involve us in their interaction. Robots that would never involve humans, even tangentially, are not relevant to discussions about integration.

At any point where humans and robots intersect, we have the power to refuse to play the part a developer's multi-agent model has imagined for us.**The point of failure is the robot's behaviour soliciting social recognition, and the skeptic's refusal to reciprocate this request.** Refusing to defer to their right of way is our most documented example, but any interaction with an element of mutual recognition is vulnerable to refusal.

One reason we, as a skeptic, may not accept the robot's invitation to treat it as a social agent is because we do not recognize in it a fellow member of society. **This is because the skeptic perceives their own agency as different from the robot.** Even if the robot mimics the correct social signals, we cannot change a sincere lack of belief in the authenticity of the exchange.

Where does our sincere lack of belief come from, then? The robot was designed to be autonomous, after all, so if Turing is to be believed then we should only require a convincing social imitator. The existence of Searle, however, shows that not everyone focuses on external expressions to judge agency.**In order to be recognizably autonomous, the robot would have to be intelligent and independent.** On these terms both theory and counter-theory agree, the problem is whether we perceive independence within the robot.

As a skeptic, we do not, because behind the interaction in front of us we do not see the intentions of a self-determined robot agent, but the one who built them. **The robot is not independent, its actions are motivated by the desires of owners and developers, to whom the robot is merely a tool through which they interact with society.** All of the forces behind the products described as imminent in our first chapter are commercial, profit-seeking ventures. It is in the interest of such ventures that other people socially recognize and reciprocate with artificial social agents where it might benefit the owner through an empowered robot, but without imposing corresponding social responsibilities on owners toward the robot. Essentially, convincing the public that their robot is one of Dautenhan's organic agents [24], while privately treating them as an artificial agent.

But the public is wise to this sort of sleight of hand, especially from commercial quarters. While businesses may already represent an ecosystem of artificial agents [59], recent attempts to put a relatable face on those corporations through social media outreach have triggered backlash not unlike Agency Alienation in character[1]. Meanwhile in research, it is an implicit premise of "Wizard of Oz" studies that the participant's ignorance of how the robot really works helps create a useful suspension of disbelief.

When push comes to shove, as when we considered the moral agency of a self-driving car, the illusion of the robot as a moral agent in its own right is quickly discarded by most of the public [12]. Other studies admit this awareness in the adult population by employing children, as in one cast that compared introducing different children to a robot as either a friend or a tool to study the effect of framing on their behaviour [115]. It is this tension between the social nature of the interaction as roboticists try to frame it and the subject's knowledge of their real motivations that blocks their ability to sustain belief in the robot's intentionality.

Having explored this hypothetical reversed perspective, it is our original theory that now appears upside down - where we decided we want the fruits of a socially-recognized autonomous robot for our own benefit, and are working backward from that point to achieve them, rather than trying to invent a truly independent social agent first and then seeing what such agents might achieve. The problem lies not with the public failing to buy in to the robot's Social Agency, but with the tainted origin of that agency itself, disconnected from anything recognizable as personal autonomy.

[1] https://www.vulture.com/2019/06/brand-twitter-jokes-history.html

Can true Social Agency even be developed under these conditions, then? Should it, if it will be used as a means to exploit the customs of society for the benefit of owners over the broader public? How can the root cause of Agency Alienation be healed, rather than merely suppressed? For the conscientious roboticist, who recognizes the phenomena, arguments, and theories put forward by this thesis, these are the next great questions that demand investigating.

6.2 Additional Implications

Setting aside this central argument concerning the nature of Social Agency, there are a number of secondary effects and consequences we may also observe, any one of which could prove fruitful to future researchers.

6.2.1 Superhuman Social Agents

Autonomous robots embedded in human society offer a unique opportunity for hand-crafting social agents that are not limited by human nature. Whole swaths of game theory, sociology and psychology explore the boundaries of what can be expected of humans, identifying problems like “the tragedy of the commons” where the cumulative result of each person’s rational decision unintentionally destroys a common good. Games like “the prisoner’s dilemma” are meant to capture exploits and vulnerabilities in human thinking difficult or even impossible to overcome.

Robots are not bound by any of these rules. They can place the needs of the group over personal interest, endure insults with unlimited dignity, and put their complete trust in others. This is not a purely hypothetical opportunity - in *Do You Want Your Autonomous Car To Drive Like You?* [8], the authors found many drivers wanted their autonomous cars to drive defensively, even if they themselves were aggressive drivers. Creating social agents necessarily means grappling with the quality of their conduct, and may present opportunities to raise the bar across society.

6.2.2 The Effect on Ourselves of Denying Agency to Others

Creating artificial social agents whom we only selectively recognize can have a greater impact on ourselves beyond how our doorway interactions will resolve. In an article from Slate, *I Judge Men Based On How They Talk to The Amazon Echo’s Alexa*[2], the author describes feelings of discomfort she experiences witnessing a date verbally mistreating Alexa, a virtual representation of a woman. It is a common piece of advice that when getting to know someone, pay attention to how they treat service staff for a glimpse of how they treat

[2] https://slate.com/technology/2018/04/i-judge-men-based-on-how-they-talk-to-the-amazon-echos-alexa.html

people over whom they have power, but not everyone feels this same sympathy engage for products designed to be ordered around. That many of these virtual assistants are coded as female by default - Alexa, Siri, Cortana - could also reinforce negative stereotypes of female subservience.

Becoming accustomed to autonomous robots taking on ever more complicated tasks for us carries the danger of acclimatizing us to a master-servant mentality, where the goal is explicitly to make the servant as human-like as possible without creating any duties or responsibilities in kind. We have already explored the tension between commercial motivations and the gravity of adding new members of society, but even when we are untroubled by integrating artificial social agents into our lives, they may have a subtle and shaping influence on those who make use of them.

6.2.3 Categorical Differences Between Humans and Robots

The most fundamental challenge presented by Social Agency is whether a robot - as a synthetic, inorganic entity - can ever truly be an equal to a human. Even more significant than the tension created by a developer's motives, if a categorical difference separates humans and robots from the start, there are serious questions whether social integration between humans and robots ever could (or should) succeed.

We have already mentioned the famous Turing Test and the Chinese Room Puzzle, which are ultimately different intuitions surrounding the same basic circumstances. There is no more hard evidence to solve the disagreement about the nature of artificial consciousness than there is to disprove solipsism's claim that one can only be sure of one's own consciousness. This debate has raged for thousands of years previous to the advent of robots, and a definitive answer seems unlikely.

Addressing this question would not even presume full equality between humans and robots. Animals occupy a variety of levels of recognition in human society, from laws protecting pets from abuse to the preservation of endangered species to the extermination of nuisance insects. Whether a robot can or should occupy any space in this hierarchy is a question of fundamental significance to humanity's self-perception, and it would be naive of roboticists to think they can produce work that will challenge the great mysteries of life and be left alone indefinitely.

6.3 Future Work

Beyond some of the higher theoretical implications our theory has for autonomous robotics, it also opens up further avenues of research. Reshaping the social relationship between humans and robots creates new circumstances and requirements, many with practical applications.

6.3.1 Common Channels of Communication

Managing a robot's Social Agency in part means accepting the importance of operating transparently and in the open. Sensing and communicating are socially sensitive actions among humans even when not directly aimed at each other, and the same will likely hold for robots attempting to integrate with them. One example of a cocktail party robot we discussed while introducing the incidental interface captured this sort of mindset - humans at a cocktail party would like to use their voice to signal their desire for a drink, so a robot waiter should be sensitive to this signal [25]. There are other ways a robot could localize a human or identify their desire for a drink, but using a common communication channel allows for more human-like social engagement than a more opaque medium.

Our own incidental interface project explored how this can even come into play in robot-robot interactions around humans. Robots may be able to communicate wirelessly, but network messaging can leave humans out of the conversation or make the sudden reactions of robots seem arbitrary or unprompted, as though overhearing an argument in a foreign language. Passive social signals for mood or urgency can let bystanders understand a social situation they may not be directly involved in, and warn them of when to intervene or steer clear. Participating in these layers with humans can be one way of living up to the social expectations that agency creates.

6.3.2 The Perception of Learning vs. Hand-Crafted Systems

For some participants of the doorway interaction studies, particularly those who changed their behaviour over the course of their interactions, hearing an explanation of how the behavior worked may have turned it into a mere mechanism rather than a social interaction. Several chose to play with the robot, deliberately triggering a reaction over and over to see if it would respond as predicted.

Knowing that a system and its behaviours are the product of a developer's design decisions rather than the robot's own "free will" may color users' perceptions by making it predictable, another facet of intentionality. This again is the implicit assumption in many "Wizard of Oz" style studies, where the robot's behaviour is to some degree an illusion that would be spoiled if the participant was informed of how it really worked.

One interesting possibility is that behaviours that are learned rather than hand-crafted might appear more authentic, even if there were no outward difference between the two. Putting even one layer of abstraction between the developer's intentions and what the robot learns creates space to imagine the two as separate entities, even if one is heavily controlled and shaped by the other. Neural network based approaches in particular, by virtue of having some relation to biological intelligence and learning, may carry greater weight with the public's sense of what constitutes intent, something that could be investigated with future studies.

6.3.3 Measuring Social Buy-In

The development arc of this thesis made heavy use of experiments that could isolate and measure the effect that social recognition by humans of robots was having on their successful integration with one another. To that end, the concept of “social buy-in” - that is, whether participants bought into their encounter with the robot as a genuine “social interaction” - was a leading metric when discussing concepts like right of way.

Social buy-in should be distinguished from merely having the interaction outwardly completing as intended. A robot might be able to assert itself in navigation using brute force to cause the human to give way, but this should not be mistaken for deferring to the robot’s right, as we saw in several cases with the assured behaviour. Social recognition and reciprocation takes place by the consent of the user, so attempts to cultivate it in studies must track both the participant’s behaviours and their beliefs, and correlate which actions affect which.

6.4 Limitations

While concluding the business of this thesis, it is important to acknowledge some of the limits of the research these larger theories and principles were built on. Our study of doorways, for example, acknowledged our exclusive focus on the “base case” of two parties approaching from opposite sides - expanding to consider more complicated interactions, such as multiple parties or starting from the same side, would make that particular product more robust and useful, but was beyond our scope. Conversely, moving away from doorways entirely to new situations would provide more opportunities to test for the phenomena that we have attributed to Social Agency. Other navigation challenges like corridors or crowds have their own social customs, and Agency Alienation may express itself in different ways.

Further studies with larger or more diverse pools of participants are needed to cement our findings, which despite being drawn over three studies, still depend on a relatively small number of participants. This could help account for specific cultural factors that differ between populations [6]. This is a criticism that holds for all three projects herein, but also the field of human-robot interaction overall. Should the predicted new wave of autonomous robots begin making their appearance and interacting with the population at large, this problem of limited data may become moot, though that may also be too late to try and shape the public’s reaction.

Of course, improvements to the chosen platform and expanded or improved capabilities might always generate higher performance results. Our emphasis on simple robots, sensing, and means of interaction were meant to keep the focus on the impact of the robot’s own behaviours rather than overwhelming participants with impressive hardware and effects, but other researchers seeking the limits of what’s possible may want to employ more advanced humanoid robots and engaging methods like speech and eye contact.

6.5 Conclusion

This thesis articulated the concept of Social Agency as a key emerging issue on the eve of a vast effort to integrate autonomous robots into human environments. It made this case both through theory and deduction via existing literature, and through an arc of development that took previous autonomous robot systems and redesigned them for new shared spaces.

Over the course of this development arc we produced a number of useful autonomous systems for both robot-robot and human-robot interaction, including the incidental interface, the assertive behaviour, and the assured behaviour. By exploring first task allocation and supervision and then doorway navigation, we refined our focus toward the outsized influence that human social recognition played in the success of these systems. Successive studies that built off of one another both helped identify Social Agency and acted as an example of how to cultivate it through cycles of development.

The results of these investigations included linking the concepts of social recognition, reciprocation, and performance, so that pursuing the acknowledgement of a robot's Social Agency becomes a means of pursuing its effectiveness as an autonomous agent. It also presented Agency Alienation, which poses a potential danger to the project of human and robot integration, when attempts to create socially sensitive and engaging robots instead inspire backlash.

Most of all, this investigation encourages more serious consideration of the consequences of sending a new generation of autonomous robots deeper into public life. Whether the self-driving car, the delivery drone, the robotic companion, or the most fanciful imaginings of science fiction, the full implications of seizing the power to create new pseudo-members of society demands more from roboticists, both as diligent engineers, and as moral and philosophical actors.

Bibliography

[1]

[2] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 961–971, 2016.

[3] A. Badali, J. M. Valin, F. Michaud, and P. Aarabi. Evaluating real-time audio localization algorithms for artificial audition in robotics. In *Intelligent Robots and Systems*, pages 2033–2038, October 2009.

[4] Ritta Baddoura and Gentiane Venture. Social vs. useful hri: experiencing the familiar, perceiving the robot as a sociable partner and responding to its actions. *International Journal of Social Robotics*, 5(4):529–547, 2013.

[5] Brian P Bailey, Joseph A Konstan, and John V Carlis. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Interact*, volume 1, pages 593–601, 2001.

[6] Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura. The influence of peopleâĂŹs culture and prior experiences with aibo on their attitude towards robots. *Ai & Society*, 21(1-2):217–230, 2007.

[7] Christoph Bartneck, Michel Van Der Hoek, Omar Mubin, and Abdullah Al Mahmud. Daisy, daisy, give me your answer do!: switching off a robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 217–222. ACM, 2007.

[8] Chandrayee Basu, Qian Yang, David Hungerman, Mukesh Singhal, and Anca D Dragan. Do you want your autonomous car to drive like you? In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 417–425. ACM, 2017.

[9] George A Bekey. *Autonomous robots: from biological inspiration to implementation and control.* MIT Press, 2005.

[10] Cindy L Bethel, Deborah Eakin, Sujan Anreddy, James Kaleb Stuart, and Daniel Carruth. Eyewitnesses are misled by human but not robot interviewers. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, pages 25–32. IEEE, 2013.

[11] Cindy L Bethel and Robin R Murphy. Affective expression in appearance constrained robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 327–328. ACM, 2006.

[12] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars. *Proceedings of the IEEE*, 107(3):502–504, 2019.

[13] Jürgen Brandstetter, Péter Rácz, Clay Beckner, Eduardo B Sandoval, Jennifer Hay, and Christoph Bartneck. A peer pressure experiment: Recreation of the asch conformity experiment with robots. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 1335–1340. IEEE, 2014.

[14] Cynthia Breazeal. Toward sociable robots. *Robotics and autonomous systems*, 42(3-4):167–175, 2003.

[15] Cynthia Breazeal. Social interactions in hri: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2):181–186, 2004.

[16] Cynthia Breazeal and Brian Scassellati. How to build robots that make friends and influence people. In *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No. 99CH36289)*, volume 2, pages 858–863. IEEE, 1999.

[17] Stephen A Brewster, Peter C Wright, and Alastair DN Edwards. A detailed investigation into the effectiveness of earcons. In *Santa Fe Institute Studies In The Sciences of Complexity-Proceedings*, volume 18, pages 471–471. Addison-Wesley Publishing Co, 1994.

[18] Rodney A Brooks. Elephants don't play chess. *Robotics and autonomous systems*, 6(1-2):3–15, 1990.

[19] E. Cha, A. Dragan, J. Forlizzi, and S. Srinivasa. Effects of speech on perceived capability. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 134–135, March 2014.

[20] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. Socially aware motion planning with deep reinforcement learning. *arXiv preprint arXiv:1703.08862*, 2017.

[21] Kari Gwen Coleman. Android arete: Toward a virtue ethic for computational agents. *Ethics and Information Technology*, 3(4):247–265, 2001.

[22] Nikolaus Correll, Kostas E Bekris, Dmitry Berenson, Oliver Brock, Albert Causo, Kris Hauser, Kei Okada, Alberto Rodriguez, Joseph M Romano, and Peter R Wurman. Analysis and observations from the first amazon picking challenge. *IEEE Transactions on Automation Science and Engineering*, 15(1):172–188, 2016.

[23] Mike Daily, Youngkwan Cho, Kevin Martin, and Dave Payton. World embedded interfaces for human-robot interaction. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, pages 6–pp. IEEE, 2003.

[24] Kerstin Dautenhahn. Ants donâĂŹt have friends–thoughts on socially intelligent agents. *Socially intelligent agents*, pages 22–27, 1997.

[25] Antoine Deleforge and Radu Horaud. The cocktail party robot: Sound source separation and localisation with an active binaural head. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 431–438. ACM, 2012.

[26] Amol Deshmukh and Ruth Aylett. Exploring socially intelligent recharge behaviour for human-robot interaction. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 150–151. IEEE, 2014.

[27] Amol Deshmukh, Mei Yii Lim, Michael Kriegel, Ruth Aylett, Kyron Du Casse, Kerstin Dautenhahn, et al. Managing social constraints on recharge behaviour for robot companions using memory. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 129–130. IEEE, 2011.

[28] Anca D Dragan, Rachel M Holladay, and Siddhartha S Srinivasa. An analysis of deceptive robot motion. In *Robotics: science and systems*, page 10. Citeseer, 2014.

[29] Vanessa Evers, Roelof de Vries, and Paulo Alvito. Designing interruptive behaviors of a public environmental monitoring robot. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 131–132. ACM, 2011.

[30] Stephen M Fiore, Norman L Badler, Lotzi Boloni, Michael A Goodrich, Annie S Wu, and Jessie Chen. Human-robot teams collaborating socially, organizationally, and culturally. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 465–469. SAGE Publications Sage CA: Los Angeles, CA, 2011.

[31] Stephen M Fiore, Travis J Wiltshire, Emilio JC Lobato, Florian G Jentsch, Wesley H Huang, and Benjamin Axelrod. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in psychology*, 4:859, 2013.

[32] Paolo Fiorini and Zvi Shiller. Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research*, 17(7):760–772, 1998.

[33] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997.

[34] Marlena R Fraune, Satoru Kawakami, Selma Sabanovic, P Ravindra S De Silva, and Michio Okada. Three's company, or a crowd?: The effects of robot number and behavior on hri in japan and the usa. In *Robotics: Science and Systems*, 2015.

[35] Marlena R Fraune, Steven Sherrin, Selma Sabanović, and Eliot R Smith. Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 109–116. ACM, 2015.

[36] William W Gaver, Randall B Smith, and Tim O'Shea. Effective sounds in complex systems: The arkola simulation. In *CHI*, volume 91, pages 85–90, 1991.

[37] Brian P Gerkey and Maja J Matarić. A formal analysis and taxonomy of task allocation in multi-robot systems. *The International Journal of Robotics Research*, 23(9):939–954, 2004.

[38] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C Schultz, et al. Designing robots for long-term social interaction. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1338–1343. IEEE, 2005.

[39] Matthew C Gombolay, Reymundo A Gutierrez, Shanelle G Clarke, Giancarlo F Sturla, and Julie A Shah. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, 39(3):293–312, 2015.

[40] Anders Green and Helge Hüttenrauch. Making a case for spatial prompting in human-robot communication. In *Workshop Programme*, volume 10, page 52, 2006.

[41] Victoria Groom, Jimmy Chen, Theresa Johnson, F Arda Kara, and Clifford Nass. Critic, compatriot, or chump?: Responses to robot blame attribution. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 211–218. IEEE Press, 2010.

[42] Stephen J Guy, Jatin Chhugani, Sean Curtis, Pradeep Dubey, Ming Lin, and Dinesh Manocha. Pledestrians: a least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics symposium on computer animation*, pages 119–128. Eurographics Association, 2010.

[43] Jerome Guzzi, Alessandro Giusti, Luca M Gambardella, and Gianni A Di Caro. Local reactive robot navigation: A comparison between reciprocal velocity obstacle variants and human-like behavior. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2622–2629. IEEE, 2013.

[44] Donna Haraway. A cyborg manifesto. *New York*, page 150, 1991.

[45] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[46] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[47] Juan Camilo Gamboa Higuera and Gregory Dudek. Fair subdivision of multi-robot tasks. In *2013 IEEE International Conference on Robotics and Automation*, pages 3014–3019. IEEE, 2013.

[48] Juan Camilo Gamboa Higuera, Anqi Xu, Florian Shkurti, and Gregory Dudek. Socially-driven collective path planning for robot missions. In *2012 Ninth Conference on Computer and Robot Vision*, pages 417–424. IEEE, 2012.

[49] Owen Holland, Chris Melhuish, and Steve Hoddell. Chorusing and controlled clustering for minimal mobile agents. In *Proc. European Conference on Artificial Life*. Citeseer, 1997.

[50] Jim Johnson. Mixing humans and nonhumans together: The sociology of a door-closer. *Social problems*, 35(3):298–310, 1988.

[51] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Brian T Gill, Jolina H Ruckert, Solace Shen, Heather E Gary, Aimee L Reichert, Nathan G Freier, and Rachel L Severson. Do people hold a humanoid robot morally accountable for the harm it causes? In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 33–40. ACM, 2012.

[52] Ioannis Karamouzas, Peter Heil, Pascal van Beek, and Mark H Overmars. A predictive collision avoidance model for pedestrian simulation. *MIG*, 9:41–52, 2009.

[53] Harmish Khambhaita and Rachid Alami. Viewing robot navigation in human environment as a cooperative activity. *arXiv preprint arXiv:1708.01267*, 2017.

[54] Beomjoon Kim and Joelle Pineau. Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics*, 8(1):51–66, 2016.

[55] Nathan Kirchner, Alen Alempijevic, and Gamini Dissanayake. Nonverbal robot-group interaction using an imitated gaze cue. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 497–504. IEEE, 2011.

[56] Takuya Kitade, Satoru Satake, Takayuki Kanda, and Michita Imai. Understanding suitable locations for waiting. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 57–64. IEEE Press, 2013.

[57] G Ayorkor Korsah, Anthony Stentz, and M Bernardine Dias. A comprehensive taxonomy for multi-robot task allocation. *The International Journal of Robotics Research*, 32(12):1495–1512, 2013.

[58] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[59] Benjamin Kuipers. An existing, ecologically-successful genus of collectively intelligent artificial creatures. *arXiv preprint arXiv:1204.4116*, 2012.

[60] Rainer Kümmerle, Michael Ruhnke, Bastian Steder, Cyrill Stachniss, and Wolfram Burgard. A navigation system for robots operating in crowded urban environments. In *2013 IEEE International Conference on Robotics and Automation*, pages 3225–3232. IEEE, 2013.

[61] Jonathan RT Lawton, Randal W Beard, and Brett J Young. A decentralized approach to formation maneuvers. *IEEE transactions on robotics and automation*, 19(6):933–941, 2003.

[62] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. Personalization in hri: A longitudinal field experiment. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 319–326. IEEE, 2012.

[63] Dan Leyzberg, Eleanor Avrunin, Jenny Liu, and Brian Scassellati. Robots that express emotion elicit better human teaching. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 347–354. ACM, 2011.

[64] Alexandru Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati. Evidence that robots trigger a cheating detector in humans. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 165–172. ACM, 2015.

[65] Yuri K Lopes, Stefan M Trenkwalder, André B Leal, Tony J Dodd, and Roderich Groß. Supervisory control theory applied to swarm robotics. *Swarm Intelligence*, 10(1):65–97, 2016.

[66] Bertram F Malle, Matthias Scheutz, Jodi Forlizzi, and John Voiklis. Which robot am i thinking about?: The impact of action and appearance on people's evaluations of a moral robot. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 125–132. IEEE Press, 2016.

[67] Richard E Mayer, Kristina Sobko, and Patricia D Mautone. Social cues in multimedia learning: Role of speaker's voice. *Journal of educational Psychology*, 95(2):419, 2003.

[68] David McFarland and Emmet Spier. Basic cycles, utility and opportunism in self-sufficient robots. *Robotics and Autonomous Systems*, 20(2-4):179–190, 1997.

[69] J. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt. Speaking swarmish: Human-robot interface design for large swarms of autonomous mobile robots. In *AAAI Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before*, pages 72–75, March 2006.

[70] Wim Meeussen, Eitan Marder-Eppstein, Kevin Watts, and Brian P Gerkey. Long term autonomy in office environments. In *ALONE Workshop, In Proceedings of Robotics: Science and Systems (RSSâĂŹ11), Los Angeles, USA*, 2011.

[71] Wim Meeussen, Melonee Wise, Stuart Glaser, Sachin Chitta, Conor McGann, Patrick Mihelich, Eitan Marder-Eppstein, Marius Muja, Victor Eruhimov, Tully Foote, et al. Autonomous door opening and plugging in with a personal robot. In *2010 IEEE International Conference on Robotics and Automation*, pages 729–736. IEEE, 2010.

[72] Marek P Michalowski, Carl DiSalvo, Didac Busquets, Laura M Hiatt, Nik A Melchior, Reid Simmons, and Selma Sabanovic. Socially distributed perception. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 349–350. ACM, 2006.

[73] Stanley Miligram. Obedience to authority: An experimental view. *New York*, 1974.

[74] Masahiro Mori, Kari F MacDorman, and Norri Kageki. The uncanny valley: The original essay by masahiro mori. *IEEE Spectrum*, pages 98–100, 2012.

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

[76] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 61–68. ACM, 2009.

[77] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 69–76. ACM, 2009.

[78] Yasushi Nakauchi and Reid Simmons. A social robot that stands in line. *Autonomous Robots*, 12(3):313–324, 2002.

[79] Clifford Nass, Youngme Moon, Brian J Fogg, Byron Reeves, and D Christopher Dryer. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239, 1995.

[80] Elena Pacchierotti, Henrik I Christensen, and Patric Jensfelt. Evaluation of passing distance for social robots. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 315–320. IEEE, 2006.

[81] 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.

[82] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, Tonghui Zhu, and Cynthia Breazeal. Backchannel opportunity prediction for social robot listeners. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 2308–2314. IEEE, 2017.

[83] Sachin Patil, Jur Van Den Berg, Sean Curtis, Ming C Lin, and Dinesh Manocha. Directing crowd simulations using navigation fields. *IEEE transactions on visualization and computer graphics*, 17(2):244–254, 2011.

[84] A. Pazhitnov, V. Gerasimov, and D. Pavlovsky. Tetris. *Spectrum Holobyte*, 1995.

[85] Rolf Pfeifer. Building fungus eaters: Design principles of autonomous agents. *From animals to animats*, 4:3–12, 1996.

[86] 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.

[87] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3.

[88] Maria Ralph and Medhat A Moussa. On the effect of the user's background on communicating grasping commands. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 353–354. ACM, 2006.

[89] 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.

[90] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. In *ACM SIGGRAPH computer graphics*, volume 21.

[91] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *European conference on computer vision*, pages 549–565. Springer, 2016.

[92] Stephanie Rosenthal and Manuela Veloso. Using symbiotic relationships with humans to help robots overcome limitations. In *Workshop for Collaborative Human/AI Control for Interactive Experiences*, 2010.

[93] Maaike Roubroeks, Jaap Ham, and Cees Midden. When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, 3(2):155–165, 2011.

[94] Seyed Abbas Sadat and Richard T Vaughan. Bravo: Biased reciprocal velocity obstacles break symmetry in dense robot populations. In *2012 Ninth Conference on Computer and Robot Vision*, pages 441–447. IEEE, 2012.

[95] Martin Saerbeck and Christoph Bartneck. Perception of affect elicited by robot motion. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 53–60. IEEE, 2010.

[96] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to approach humans?: strategies for social robots to initiate interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 109–116. ACM, 2009.

[97] Paul Saulnier, Ehud Sharlin, and Saul Greenberg. Exploring minimal nonverbal interruption in hri. In *RO-MAN, 2011 IEEE*, pages 79–86. IEEE, 2011.

[98] John R Searle. Minds, brains, and programs. *Behavioral and brain sciences*, 3(3):417–424, 1980.

[99] Onn Shehory and Sarit Kraus. Methods for task allocation via agent coalition formation. *Artificial intelligence*, 101(1-2):165–200, 1998.

[100] Masahiro Shiomi, Francesco Zanlungo, Kotaro Hayashi, and Takayuki Kanda. Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *International Journal of Social Robotics*, 6(3):443–455, 2014.

[101] M Silverman, Boyoon Jung, Dan Nies, and G Sukhatme. Staying alive longer: Autonomous robot recharging put to the test. *Center for Robotics and Embedded Systems (CRES) Technical Report CRES*, 3:015, 2003.

[102] Tim Smithers. Autonomy in robots and other agents. *Brain and Cognition*, 34(1):88–106, 1997.

[103] Luc Steels. When are robots intelligent autonomous agents? *Robotics and Autonomous systems*, 15(1-2):3–9, 1995.

[104] Leila Takayama, Wendy Ju, and Clifford Nass. Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 25–32. ACM, 2008.

[105] Jack Thomas and Richard Vaughan. After you: Doorway negotiation for human-robot and robot-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3387–3394. IEEE, 2018.

[106] Cristen Torrey, Susan Fussell, and Sara Kiesler. How a robot should give advice. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 275–282. IEEE Press, 2013.

[107] Ali E Turgut, Hande Çelikkanat, Fatih Gökçe, and Erol Şahin. Self-organized flocking in mobile robot swarms. *Swarm Intelligence*, 2(2-4):97–120, 2008.

[108] Alan M Turing. Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer, 2009.

[109] J. M. Valin, J. Rouat, and F. Michaud. Enhanced robot audition based on microphone array source separation with post-filter. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2123–2128, September 2004.

[110] Jur Van den Berg, Ming Lin, and Dinesh Manocha. Reciprocal velocity obstacles for real-time multi-agent navigation. In *2008 IEEE International Conference on Robotics and Automation*, pages 1928–1935. IEEE, 2008.

[111] Richard Vaughan. Massively multi-robot simulation in stage. *Swarm intelligence*, 2(2-4):189–208, 2008.

[112] Lin Wang, Pei-Luen Patrick Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. When in rome: the role of culture & context in adherence to robot recommendations. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 359–366. IEEE Press, 2010.

[113] Jens Wawerla and Richard T Vaughan. A fast and frugal method for team-task allocation in a multi-robot transportation system. In *2010 IEEE International Conference on Robotics and Automation*, pages 1432–1437. IEEE, 2010.

[114] Astrid Weiss, Judith Igelsböck, Manfred Tscheligi, Andrea Bauer, Kolja Kühnlenz, Dirk Wollherr, and Martin Buss. Robots asking for directionsâĂŤthe willingness of passers-by to support robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 23–30. IEEE, 2010.

[115] Jacqueline M Kory Westlund, Marayna Martinez, Maryam Archie, Madhurima Das, and Cynthia Breazeal. Effects of framing a robot as a social agent or as a machine on children's social behavior. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 688–693. IEEE, 2016.

[116] Anqi Xu and Gregory Dudek. Trust-driven interactive visual navigation for autonomous robots. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3922–3929. IEEE, 2012.

[117] Anqi Xu and Gregory Dudek. Maintaining efficient collaboration with trust-seeking robots. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 3312–3319. IEEE, 2016.

[118] Zhi Yan, Nicolas Jouandeau, and Arab Ali Cherif. A survey and analysis of multi-robot coordination. *International Journal of Advanced Robotic Systems*, 10(12):399, 2013.

[119] Sangseok You, Jiaqi Nie, Kiseul Suh, and S Shyam Sundar. When the robot criticizes you...: self-serving bias in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 295–296. ACM, 2011.

[120] Fang Yuan, Lukas Twardon, and Marc Hanheide. Dynamic path planning adopting human navigation strategies for a domestic mobile robot. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3275–3281. IEEE, 2010.

[121] Yinan Zhang and Richard Vaughan. Ganging up: Team-based aggression expands the population/performance envelope in a multi-robot system. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 589–594. IEEE, 2006.

[122] Mauricio Zuluaga and Richard Vaughan. Reducing spatial interference in robot teams by local-investment aggression. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2798–2805. IEEE, 2005.