FAKULTÄT FÜR
MASCHINENWESEN UND
BETRIEBSWISSENSCHAFTEN
Mechanical & Industrial Engineering

Designing for Interactions That Enable End User Impacts on Robot Behavior

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Dr.rer.soc.oec.

by

MSc Helena Anna Frijns
Registration Number 11832418

to the Faculty of Mechanical and Industrial Engineering

at the TU Wien

Advisor: Univ.-Prof. Mag. Dr. Sabine Theresia Köszegi

The dissertation has been reviewed by:

Assistant Prof. Dr.in phil.
Mag.a phil. Astrid Weiss

Prof. Dr.-Ing. Anna-Lisa
Vollmer

Vienna, 14th May, 2024

Helena Anna Frijns

Declaration of Authorship

MSc Helena Anna Frijns

I declare in lieu of oath, that I wrote this thesis and carried out the associated research myself, using only the literature cited in this volume. If text passages from sources are used literally, they are marked as such. I confirm that this work is original and has not been submitted for examination elsewhere, nor is it currently under consideration for a thesis elsewhere. I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright.

Vienna, 14th May, 2024

Helena Anna Frijns

Acknowledgements

The work conducted for this dissertation was funded by TU Wien, Doctoral College TrustRobots, as well as an FWF #ConnectingMinds grant to the project CM 100-N, Caring Robots // Robotic Care.

I would especially like to thank the supervisors of my dissertation, Dr. Oliver Schürer and Prof. Sabine Theresia Köszegi. Thanks go out to my colleagues at the Institute of Management Science, research group Labor Science and Organization. Thanks to my colleagues in the project Caring Robots // Robotic Care, Sabine, Markus, Astrid, Margrit, Chris, Doris, Carina, Nicole, Evelyn, Jürgen, Ralf, Darja, Matthias, Reinhard, Anna, Laura, and Katharina. Thanks to the co-authors of the publications that are part of this dissertation and supporting publications. Thanks go out to all study participants.

I would like to thank my colleagues and friends that were part of the Doctoral College TrustRobots. This dissertation would not be the same without all the discussions and collaborations we had over the years, where we were able to learn from each other. With Darja Stoeva I worked with dancers (with Oliver and Eva in H.A.U.S.), and developed an imitation system, which was the foundation of one of the papers in this dissertation. I collaborated with Christina Schmidbauer on a paper on design guidelines for cobot programming interfaces. With Matthias Hirschmanner, I explored the topics of transparency in HRI and human-in-the-loop error detection, writing a paper on the design of interfaces for those purposes. From Glenda Hannibal I learned more about the role of theory in HRI. With Florian Beck, Darja and I discussed robot kinematics. Michael Koller worked on the topic of joint attention, which is closely linked to the joint action topic that I explore here. Setareh Zafari's perspectives on machine agency form a counterweight to my perspective on supporting human agency in human-machine interactions. Guglielmo Papagni worked on the topic of explainability, which is strongly linked to my interest in making it possible for end users to interpret robot behavior. From Jesse de Pagter, I learned about importance of existing cultural narratives on choices that are and will be made regarding governance and design of robotics and AI. Isabel Schwaninger worked on HRI and HCI research in the care context, prior to my own explorations in that field. With Dominik Bauer's work, it was interesting to see the connection between the technical capability of pose estimation and representations that are explainable and that can visually be interpreted. I would like to express my gratitude to the main organizers of this doctoral college, Prof. Sabine Köszegi and Prof. Markus Vincze, who decided to bet on interdisciplinarity.

Thanks also to family, friends, and people along the way. Thanks Ace. Thanks go out to my sisters Saskia and Anne, and their partners Bart and Fernando. Thank you Andrea and Kami for the adventures! Thank you Matthias for supporting my goals. Finally, I wish to thank my parents, you are awesome.

Kurzfassung

Diese Dissertation befasst sich mit der Frage, wie die Interaktion mit Robotern so gestaltet werden kann, dass Endbenutzer:innen Einfluss auf das Verhalten des Roboters haben können. Ich untersuche die Kombination von Entscheidungen beim Interaktionsdesign und bei der Systemarchitektur, und wie sich diese auf die Interpretation der Technologie durch die Endbenutzer:innen auswirken und diese in die Lage versetzen, mit der Technologie zu agieren.

Warum sollte das Interaktionsdesign von Robotersystemen berücksichtigt werden? Unterschiedliche Roboter können so programmiert werden, dass sie unterschiedliche Dinge tun, selbst wenn sie die gleiche Gestalt haben. Die Gestalt eines Roboters kann bestimmte Erwartungen hinsichtlich der Fähigkeiten des Roboters hervorrufen. Zum Beispiel können humanoide Roboter Erwartungen an menschenähnliche Interaktionsmöglichkeiten hervorrufen, die der Roboter möglicherweise nicht erfüllen kann. Oft werden Kommunikationsmodelle aus der zwischenmenschlichen Kommunikation entlehnt und von anderen Forschern auf die Mensch-Roboter-Interaktion übertragen. Dabei handelt es sich um symmetrische Modelle, die den Menschen und den Roboter mit gleichen Fähigkeiten ausstatten. Ich vertrete die Ansicht, dass wir die Mensch-Roboter-Interaktion stattdessen als asymmetrisch betrachten sollten. Welche Fähigkeiten ein Roboter besitzt, sollte durch dessen physische Form und Verhaltensdesign kommuniziert werden. Dies kann z.B. dadurch geschehen, dass das Robotersystem transparent gestaltet wird, das heißt, dass das System den Endbenutzer:innen klar macht, welche Informationen das Robotersystem verarbeitet und wie das System Entscheidungen trifft.

Warum ist es entscheidend die Endbenutzer:innen einzubeziehen? Zum Zeitpunkt der Entwicklung ist es schwierig alle potenziellen Nutzungsformen vorherzusagen. Wenn Roboter auf die reale Welt treffen, werden sie mit einer Vielzahl an unvorhersehbaren Situationen konfrontiert. Daher argumentiere ich, dass Endbenutzer:innen in die Lage versetzt werden sollten, das Verhalten des Roboters zum Zeitpunkt der Nutzung kontextgerecht zu adaptieren. Ein zweiter Grund für die Einbeziehung von Endbenutzer:innen ist, dass Endbenutzer:innen über implizites Wissen verfügen (*tacit knowledge*), das sich aus dem Körper und seinen situativen Interaktionen ergibt. Aus diesem Grund müssen wir Systeme und Methoden entwickeln, die das Einbeziehen solcher Expert:innen fördern. Drittens wird die Möglichkeit, das Verhalten des Roboters anzupassen, die Handlungsfähigkeit und Selbstwirksamkeit der Endbenutzer:innen unterstützen. Damit wird ihnen ermöglicht, im Sinne ihrer Ziele zu handeln und sich selbst dazu befähigt fühlen.

In der Dissertation untersuche ich verschiedene Methoden, um (stellvertretende) Endbenutzer:innen, Stakeholder und Personen mit spezifischem Wissen in den Gestaltungsprozess des für die jeweiligen

Anwendungsgebiete relevanten Roboter-Verhaltens einzubeziehen. Bewohner:innen von Pflegeheimen und Pflegepersonal werden für eine zu entwickelnde Technologie in die Diskussion von Anwendungsszenarien im Rahmen einer partizipativen Designstudie mit einbezogen. Ich analysiere und entwickle Schnittstellen für die Endbenutzerprogrammierung in der Fertigung und für die Programmierung von Roboteranimationen in der sozialen Interaktion und im Tanz. Ich führe eine Studie über die Darstellung der Roboterkenntnisse für Endbenutzer:innen in einem Szenario zur Objektorganisation durch. Der Beitrag meiner Arbeit ist ein besseres Verständnis der Faktoren und Methoden des Interaktionsdesigns, die die menschliche Handlungsfähigkeit in HRI-Szenarien verbessern, indem sie es ermöglichen, das Roboterverhalten zu beeinflussen.

Abstract

This dissertation is focused on how to design for interactions with robots in ways that enable end users to impact robot behavior. I consider the combination of interaction design and system architecture choices, and how these affect how end users interpret the technology and are able to act on and with the technology.

Why consider interaction design of robotic systems? Different instances of robots can be programmed to do different things, even if they have the same embodiment. The embodiment of a robot may raise certain expectations regarding what the robot is capable of. For example, humanoid robots may raise expectations of humanlike interaction capabilities that the robot may not be able to meet. Often, models are borrowed from human-human communication and applied to Human-Robot Interaction by other researchers. These are symmetrical models that depict the human and the robot agent to have similar capabilities. I argue that we should instead see human-robot interaction as asymmetric, and that what the robot is (capable of) doing should be communicated through the robot's physical and behavioral design. This can be done, for instance, by making the robotic system transparent, meaning that the system makes it clear to end user(s) what information the robotic system is processing and how the system makes decisions.

Why involve end users? It is difficult to foresee potential use at design time. When robots meet the real world, they are bound to encounter all kinds of unforeseeable situations. Therefore, I argue that end users should be enabled to make the robot's behavior more context-appropriate at use time. A second reason to involve end users is that end users have tacit knowledge, which is the knowledge that arises from the body and its situated interactions. This is why we need to design systems and methods that promote the inclusion of such experts. Third, I argue that when it is possible for end users to adapt the robot's behavior, this supports their agency and self-efficacy, so that people are enabled to act in accordance with their aims and perceive themselves as capable of doing so.

In the dissertation, I investigate different methods to include (representative) end users, stakeholders and domain experts in the process of determining robot behavior, in several different contexts. Care home residents and care workers are involved in discussing application scenarios of a to-be-developed technology in a co-design study. I study and develop end-user programming interfaces in manufacturing and for programming robot animations in social interaction and dance. Finally, I conduct a study on representing a robot's knowledge base to users in an object organization scenario. The contribution of my work is a better understanding of interaction design factors and methods that enhance human agency in HRI scenarios by making it possible to impact robot behavior.

Contents

CHAPTER 1

Introduction

This chapter is partially based on the following book chapter: Frijns, H. A., Schürer, O. (2022) Design as a Practice in Human-Robot Interaction Research. Book chapter. In S. T. Köszegi, M. Vincze (Eds.), Trust in Robots (pp. 3–29). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_1

This dissertation is concerned with the question how to design for interactions with robots, in a way that makes robot behavior understandable for human interaction partners and makes it possible to impact robot behavior. Recent years have seen developments in robotics. Applications of new robotic solutions are explored in the care context (see, e.g., [40, 166]). In manufacturing, collaborative robots (cobots) are investigated and implemented that operate in close proximity with humans and interact with them [77]. Research has described implementations of robots that take on social roles, and how humans respond and adapt to robots in their social space, for instance, in households (e.g., robot vacuum cleaners [93]) and public space (e.g., delivery robots [68]). This raises questions regarding how people will (want to) interact with robots, and how robots can become part of human social spaces in ways that enable humans and robots to mutually affect each other. However, it is an open question how robots can be sufficiently adaptive to different contexts. Towards this end, it is necessary to consider how interactions with robots can be shaped in such a way that end users (with little programming expertise) are able to interpret and affect the robot's behavior. In this dissertation, I investigate how to make the capabilities of robotic systems transparent through design. Moreover, I investigate how end users can be enabled to impact robot behavior through end user programming, in the context of interaction, and through being involved in design processes of robotic technology.

Interpreting robot behavior: When encountering a robot, it is not immediately apparent how one can expect this robot to behave. Its programming may be different from a robot that looks exactly the same. Moreover, the robot's behavior may not change over time nor be adaptive to the context or to users. Humanlike interaction and embodiment designs have frequently been applied, based on the assumptions that this will make the robot more acceptable, familiar, and predictable [86]. Research has been conducted on the design and effects of humanlike embodiments [30, 120, 181] and which features make a robot appear humanlike [213]. Others have worked on developing robot capabilities for interaction between humans and robots, such as recognizing people and generating and interpreting verbal communication, nonverbal behavior, intentional actions, socially aware navigation, and affect [275]. However, there are limitations and disadvantages to purely relying on human interaction metaphors and humanlike embodiments. While there is a plethora of research investigating humanlike communicative behaviors for robotics, there is a limited theoretical understanding regarding if and how communication between people would apply to HRI. In this dissertation, I argue that robots are (at this point in time and for the foreseeable future) asymmetric to humans with regard to their perception and action capabilities. Very humanlike interactions may be familiar, thus facilitating interaction, but also raise expectations that cannot (yet) be met. Robot form alone only partially discloses the robot's task or how it can be expected to behave, as its behavior depends on its programming. I argue that humans and robots should be understood as asymmetric regarding perception, action and decision-making capabilities. To avoid mismatches between expectations and capabilities, it is important to consider how to design the interaction in a way that people are able to interpret its capabilities and current system state and processing. In other words, we need to consider how interactions can be designed for in such a way that it is transparent what kind of information the robotic system is processing. An improved understanding of robots can help people meet their own needs in interacting with the technology.

Impacting robot behavior to make it more context-appropriate and include tacit knowledge: Besides making it easier to interpret robot behavior, I investigate ways to impact robot behavior. Being able to impact robot behavior also requires that it is in some way apparent what the robot is doing, why, and how to change the robot's behavior. Thus, in this dissertation, being able to impact the robot's behavior and making it possible for end users to interpret the robot's behavior are seen as connected. Investigating how to enable end users and other stakeholders to affect robot behavior will identify and create opportunities for bringing in (tacit) knowledge of people with domain expertise (for instance, from the contexts of dance, social interaction, care, and manufacturing). Inclusion of domain experts in the process of robot behavior design can make the interaction more contextually appropriate. Alves-Oliveira et al. [11] argue for user involvement in social robot design, as not doing so can lead to wrong assumptions regarding user needs. I investigate how to integrate end users in all phases of robot behavior design.

Supporting human agency and oversight: This dissertation contributes to the question of how to translate ideals of supporting human agency and oversight [124] into practice. Agency refers to the *"capacity to make a difference"* [234, p.22] or to exercising the capacity to act [242]. Exercising agency can be understood as performance of intentional (and potentially also additional unintentional) actions [242]. Supporting human agency is taken to mean a consideration of how to design technology in a way that spaces are opened for influence; specifically, influence by the people who are affected by the technology's actions and shaping powers. This requires giving users insight into how robotic systems work and enabling them to affect robot behavior, even if they have little programming knowledge.

In Section 1.1, I describe related work. Section 1.2 contains descriptions of the relevant terms that are used in this dissertation. In Section 1.3, I outline the methodology. Section 1.4 contains summaries and research questions of the individual chapters of this dissertation. Section 1.5 outlines the contributions of the dissertation, Section 1.6 contains further discussion, and Section 1.7 lists the publications that are part of this thesis, as well as supporting publications.

1.1 Related work

Related work includes design frameworks, methods and technologies that enable end users and stakeholders to adapt systems in the contexts of development and use. This requires systems design, interaction design, making it possible for end users to adapt robot programs, and making the way robots function understandable to end users.

Several authors proposed design frameworks aimed at integrating end users in the Human-Computer Interaction (HCI) context, notably the meta-design framework for end-user development (EUD) by Fischer and Giaccardi [87]. The meta-design framework presupposes that not all situations can be foreseen at design time. Users should be supported in becoming co-designers by creating systems that are built to evolve by user actions. They contrast design time (when designers and representative end users develop systems) to use time (when systems may need to be adapted towards user needs) [87]. At design time, designers and representative end users develop systems for what they call a “world-as-imagined”. At use time (“world-as-experienced"), systems may need to be adapted towards the user’s needs. The authors discuss *underdesign* as a strategy in meta-design that involves creating environments or design spaces for end users at design time, to allow system modification at use time. To allow for this, infrastructures for participation in the socio-technical system need to be created. Responsibility is not transferred to the user, but some control is [87]. In a similar vein, Dix proposes design for appropriation, a strategy that enables technology to be used in ways unforeseen by the designer, or even subverting intentions of designers [67]. Underdesign in the meta-design framework and designing for appropriation share conceptual similarities. Both aim to create openness for end users so they can adapt the system at use time, and the artifact and its use can evolve. Strategies for meta-design include task-specific languages, hiding low-level computational details, transparent programming environments, collaboration, and supporting customization and reuse [87]. Similarly, strategies for design for appropriation include making the way the system works clear to the user, making it explicit why the system works the way it does, making reconfiguration of the system possible, and encouraging sharing and learning from the ways users appropriate the technology [67]. While more or less established strategies exist for such interactions with desktop computers, laptops, and mobile phones, this requires more work in the context of HRI. Robotic systems are embodied, potentially even distributed in space, and a variety of interaction modalities are used to establish the interaction. Robots can move around in space and have a degree of autonomy. They may or may not have screens as part of their embodiment. Therefore, I investigate what such strategies could be for HRI. The main methods I use to explore this are co-design, End-User Programming (EUP) and transparency.

Several works describe research on participatory design (PD) or co-design for HRI [14, 16, 102, 170, 221, 257]. Co-design makes it possible for stakeholders to have a say in the design of robotic technology. Co-design processes often include activities that enable participants to express tacit knowledge and facilitate mutual learning between stakeholders and researchers [170, 233]. In PD projects for robots in the care context, the application scenario is usually predetermined (e.g. [14, 204]), while in this dissertation (Chapter 3), application scenarios are co-determined with stakeholders. Moreover, both care workers and care home residents were included in ideation activities.

EUD focuses on methods that are applicable throughout the life-cycle of software systems, including EUP [56]. The research area of EUP for robotics aims to make it possible for end users to re-specify robot behavior to better support their needs [4], or for domain experts on social interaction to develop complex, contextually appropriate robot behavior for social interactions [56, 216]. In contrast to methods such as PD and human-centered design that focus on integrating end user and domain knowledge prior to implementation, systems that allow for EUD enable end users to modify robot programs both at design time and at use time [56]. Similar work that aims to make robot programming easier for end users includes work on synthesizing programs from demonstration for social HRI [216, 217] and open-source software development for social robots (e.g., [10]). In this dissertation, I investigate interaction design for EUP and its connection to design choices regarding the system architecture.

Transparency is proposed to make robotic systems understandable to end users during interaction. Related terms include understanding, explainability, interpretability, and intelligibility [246]. Theodorou et al. [273] write that so far, the topic of transparency has mainly received attention in relation to human-robot collaboration, but they argue transparency should be a general requirement that will enable end users to develop more accurate mental models of systems over the course of interaction. In HRI, transparency can take the form of conveying a robot's decision-making or inner workings to a user, so the user is better able to understand (correct) functioning of and decision-making by the system; understanding what the technology does and to what end [84, 246]. The concept of transparency can be framed in different ways in the context of robotics and artificial agents, namely as the absence of deception regarding the artificial nature of the agent, a means for reliability and error communication, or a way to convey the system's decision-making [273]. For designing transparent systems, Theodorou et al. [273] argue that communication methods should convey information that is relevant, have the appropriate level of abstraction, and contain an appropriate amount of information, which will depend on the user and the application. Schött et al. [246] describe that robots can be made transparent by design, robot actions, or external explanations. In this dissertation, I further argue for transparency as a means to support asymmetric HRI. Moreover, I implement it in designing interfaces.

The research gap addressed by this dissertation concerns the role of interaction design in creation of spaces that allow for affecting robot behavior by end users, from the perspective that asymmetry between humans and robots should be addressed. While there is existing work on EUP and transparency, my work focuses on the connection of system architecture to interaction design. My work investigates the connection between interpretation of the way systems work, system architecture considerations, and how to enable end user impacts at various stages in design processes. Interaction design forms a bridge between the way technology is implemented and the user's interpretation of it.

1.2 Description of terms

In this section, I explain what I mean by interaction design, robots and end users in the context of this dissertation.

1.2.1 Robots

The word robot has been defined as a *“programmed actuated mechanism with a degree of autonomy (...) to perform locomotion, manipulation or positioning"* [140, 3.1], where autonomy refers to the *“ability to perform intended tasks based on current state and sensing, without human intervention"* [140, 3.2]. In the context of this dissertation, the word **robot** refers to robotic systems; configurations of components (hardware and software) that achieve a coupling to the environment and enable the robot to sense and act on this environment. This configuration of components can be adapted and impacted by humans, for instance, through programming or exchanging a component. Moreover, following the definitions, robotic systems have a degree of autonomy. This means that the system can adapt its state and behavior execution, for instance, based on sensing a change in the environment.

The robotic systems under consideration in this dissertation are intended for interaction with humans in close proximity. The specific robotic platforms (specific implementations of robotic hardware and software that are, in this case, sold by manufacturers) that I worked with throughout this dissertation are the Pepper robot [253], the Franka Emika Panda cobot [106], and the UR5 [232]. Moreover, individual technical functionalities were presented as part of the co-design workshop series in Chapter 3, namely computer vision and conversational AI.

In the dissertation, I frequently refer to Human-Robot Interaction and social robotics. **Social robots**, according to Breazeal et al. [34], are designed to be capable of interacting with people in a human-centric way. Social robots can, for instance, be developed to be able to engage in verbal and nonverbal communication, have social skills [34], evoke responses from humans (e.g., anthropomorphizing the robot), or have the ability to respond to a social environment around the robot [91]. **Human-Robot Interaction (HRI)** is the core research field of this dissertation. It is a multidisciplinary research field that integrates disciplines such as HCI, robotics, psychology, philosophy, design, and more [23]. Interdisciplinary collaboration is required to achieve successful human-robot interactions but can be challenging due to differences between disciplines (in epistemologies, methods, and goals).

1.2.2 Design in HRI

Several discourses on design exist, with different epistemological origins, which range from construing design as a problem-solving activity, to design as creation, a form of reflection, or reasoning [148]. Problem-solving views of design involve judgments that are made in determining an improved state. In making these judgments, the act of designing has a normative character (see also Section 1.6.1). Such views are common in HRI. For instance, the introduction to a special issue on design in HRI contained the statement: *“In essence, design is about understanding the current state and then designing an improved future state"* [133, p.1]. See also Bartneck et al. [22], who see transforming reality as the aim of HRI designers and engineers.

Design is one of the disciplines relevant to HRI. Lupetti et al. [180] discuss the notion of designerly HRI, which refers to the work in HRI that has a design orientation, such as developing robotic prototypes and involvement with design methodologies. When discussing the design of a robotic system, we cannot treat the system in isolation. Rather, we should consider it as designed for people and for specific situations. Botero et al. [31] describe the design space as a co-constructed space that is formed by stakeholders, technologies, and social processes. Design activity takes place in this complex socio-technical context.

Several authors describe design spaces or frameworks for social robots and HRI [20, 66, 109] that include factors such as the robot's appearance, social capabilities, role, autonomy, and information exchange. According to Deng et al. [66], the design space of social robots is made up of its social role, embodiment, and communicative behavior. These factors correspond to aspects of interaction design, industrial design, and robot animation design. HRI design methods include, for instance, animation studies, 3D modeling, sketching, brainstorming, interviews, questionnaires, PD methods, focus groups, observations, and personas [180].

1.2.3 Interaction design

Interaction design, or designing *for* interaction, involves shaping the appearance and behavior of systems in response to input, as well as the quality of interaction, in a way suitable to the context, in order to improve current systems or build new ones [81, 108, 252]. Interaction design is not only concerned with designing the shape of objects, but also with intended use and how the artifact is experienced over time [224]. Interaction design for Human-Robot Interaction (HRI) involves affecting the appearance and interaction modalities of *robotic systems* and behavior in response to input (see also [108]).

Fallman [81] describes interaction design as involving combinations of activities from design practice, design exploration, and design studies. *Design practice* involves the development of products and prototypes with a design research question in mind. *Design exploration* questions the status quo, current narratives, and preconceptions, and proposes alternatives and counternarratives. It can involve critical and speculative design (for an example of speculative design of robotic domestic products, see [15]). *Design studies* develops knowledge on design research and results from design research activity [81].

User Interface (UI) design is a relevant aspect. However, in the case of interacting with co-located robots, a person interacting with a robot will use more than screens as sources of information in the interaction. For instance, the robot's entire embodiment becomes informative to the user, including such things as motor sounds. The interaction also involves, for instance, the person(s) interacting with the system, the physical and social environment, and can include external processing of information on remote servers. To give an example, consider a concept such as "intuitive use", which involves (semi-automatic) application of prior knowledge due to familiarity with similar interactions [195]. Achieving a high level of ease of use, or intuitive use, thus appeals to prior experience, familiar cultural metaphors, motor memory, etcetera. Interaction involves more than information flow between a user and a system through predefined interaction modalities. When considering interaction design, it is beneficial and necessary to consider the (target) context, diversity of users and other stakeholders, as well as ethical and social implications.

In this dissertation, design functions as a bridging discipline between the technical possibilities of the system and its users. Designing for interactions with robotic systems that empower users requires developing systems with a high ease of use and that communicate the current system state and action possibilities, in a way that is understandable for users with limited expertise in programming and robotics. My work applies what can be called “interaction design research", in which systems are developed for interaction with people and learnings are gathered from building these prototypes. The focus of the dissertation is on robot behavior specifically, not embodiment design, though consideration will be given to audiovisual feedback modalities. The aim is to demonstrate how spaces can be created in human-technology relations for interpretation and action by end users and other stakeholders by studying technical systems and design processes.

1.2.4 End users and other stakeholders

The term *user* is a core term in discourse on technology design. While the term is too narrow to adequately describe people - there is of course more to people than technology use, and there can be more to relations mediated by technology than just use relations - in this dissertation I often use the term user as it is a generally accepted term to indicate a person interacting with technology. Redström [224] discusses how labeling people users in the context of a design process assumes that they are already using what is not designed yet. In this dissertation, I design for and with “representative" end users (see also [87]), stakeholders, and other domain experts. These participants may not have programming expertise, but do have domain expertise, where *domain* refers to the intended context of use, such as a care context. As non-designers participate, there are some similarities to the approach taken in open design practices, which involves developing open products such as open source software [3]. However, my focus is not necessarily on system access for the general public (though publications are open access) but rather on inclusion of representative stakeholders and end users with domain expertise in design processes and building systems that enable them to affect robot behavior.

A key point is that interactions are situated, and humans interacting with robotic technologies are all different. User(s) and other stakeholders, the robotic system and associated technologies, and the context are interrelated. Different users have different abilities, different information needs, and different expertise.

1.3 Methodology

The main research question of this dissertation is:

Main Research Question. ***How can we design robotic technologies and develop design processes in ways that support end users in impacting robot behavior?***

To answer the research question, three sets of studies were conducted with different ways of involving end users in idea generation and (re)specification of robot behavior. The three sets of studies involved co-design, end-user development, and transparency, to investigate different methods towards supporting human agency in HRI. This involved theoretical work, practical work to build functional systems, and conducting user studies. A variety of contexts, methods, application scenarios and robotic systems was explored. The studies are related through their joint focus on creating spaces in technology design to enable end user impacts on robot behavior. In this section, I will discuss the

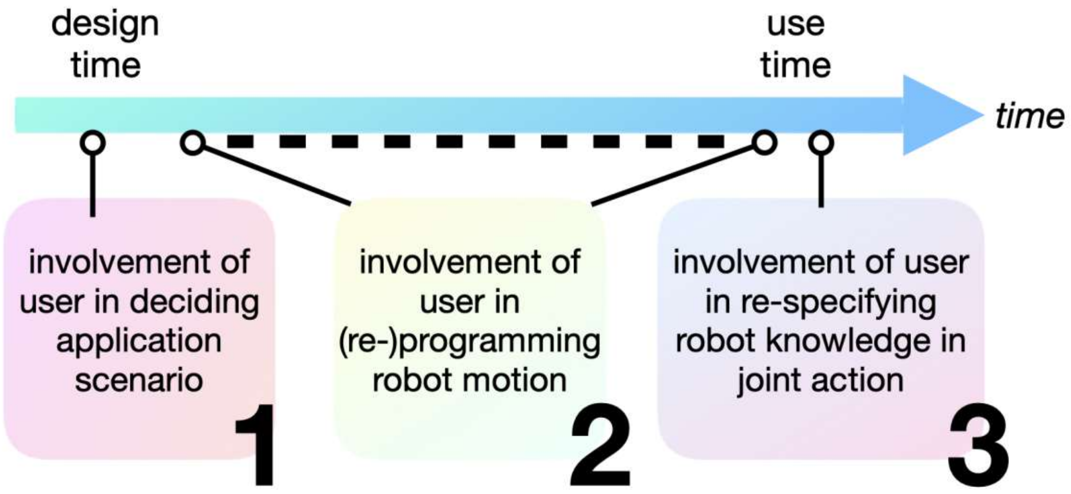

design
time
use
time
time
involvement of user in deciding application scenario
1
involvement of user in (re-)programming robot motion
2
involvement of user in re-specifying robot knowledge in joint action
3

Figure 1.1: This dissertation considers how to incorporate user impacts on robot behavior at various points in the design cycle; (1) in deciding application scenarios of robotic technology (Section 1.3.2), (2) End-User Programming (EUP) (Section 1.3.3), and (3) in the context of direct interaction (Section 1.3.4). The design time-use time continuum (top) is inspired by Fischer and Giaccardi's meta-design framework [87].

main components of the dissertation, namely theory on communication models in HRI, and practical studies on co-design, end-user development, and transparency.

To investigate forms of involvement, a variety of methods were applied at different stages of the design time-use time continuum (see Figure 1.1). The first practical study was a co-design study (Figure 1.1, (1)). Subsequent practical studies included studies on end user programming (Figure 1.1, (2)) and a study on the design of transparent interfaces (Figure 1.1, (3)). Figure 1.2 shows what types of interactions these studies involved. The co-design study (1) involved interactions between designers/developers and representative end users/stakeholders to identify potential applications of robotic technology, with the idea that the designers/developers then translate findings into development of systems. The studies on end-user programming (2) involve direct impacts of end users on the robot's program. The transparency study (3) involved direct interaction between users and a robot, including interaction via a shared representation.

In Section 1.3.1-1.3.4, I briefly explain how these studies connect to each other and the design time-use time continuum. In Section 1.4, I describe each chapter in more detail.

1.3.1 Theory on interaction

A review of existing models and understandings of communication in HRI was conducted [100] (Chapter 2). The aim was to develop a more appropriate understanding of what communication in HRI entails. Anthropomorphic design strategies are frequently applied to capitalize on the human tendency to anthropomorphize [34], with the idea that human-likeness will mean that people semi-automatically know how to interact with the system. However, there are differences between humans and robots

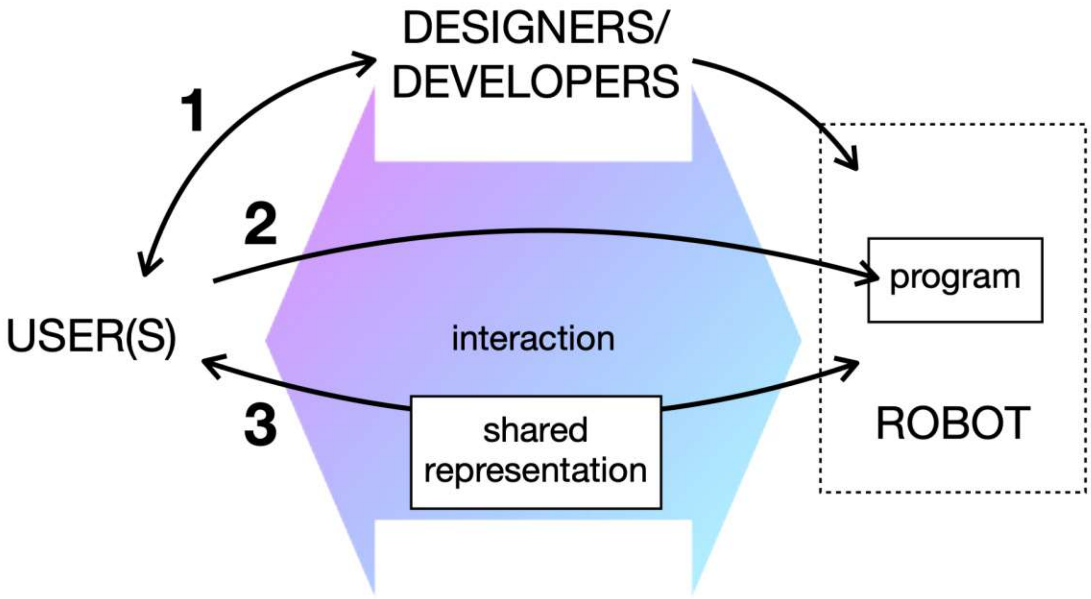


Figure 1.2: The different interaction scenarios under consideration. The dissertation considers different types of interaction, (1) interaction between users/stakeholders and designers/developers of technology, so that stakeholders can communicate preferences and are involved in determining the application scenario, (2) users directly impacting the robot program, and (3) interaction between user(s) and robot, including via a shared representation.

in terms of, e.g., perceptual processing and action capabilities. Thus, the application of human-like design metaphors requires careful consideration to prevent communication failures. A conceptual model of asymmetry between humans and robots was developed, which informed interaction design recommendations [100]. Due to this asymmetry, it is important to make robot capabilities explicit and to signify how the robot can be interacted with, for instance, through transparency. Transparency is a way to convey, e.g., robot decision-making to a user, who is then better able to understand how the system works [84, 246, 273]. Another example design recommendation is the use of shared representations to keep track of common ground between humans and robots involved in interaction.

1.3.2 Co-Design

A co-design study was conducted that involved care stakeholders (12 care home residents and 13 care workers) in the process of determining application scenarios of (robotic) technology in the care context. The aim of this study was to highlight needs that exist in the care context and to enable a process of mutual learning between technology developers and care stakeholders. The co-design process included workshops on robotic technology, conversational AI and computer vision, with technology demonstrations. Thus, it aimed to make the way technologies process information insightful for participants so that they could better reflect on potential uses of such technologies. This resulted in considerations on how to conduct a co-design process in which the application scenario and technology are fully open, with both care home residents and care workers.

1.3.3 End-User Programming

The second way of involving end users was through an approach that combined EUP and interaction design, by investigating design factors important in UI design and interaction design for EUP applications. Programming is a way of making a robot's behavior concrete and shaping how it will react in context, even if the behavior in context cannot fully be determined at design (/programming) time. EUP can, however, allow for reprogramming in the context of use. This dissertation investigates easy to use input forms and programming tools, bringing programming tools for robotic systems closer to the domain expert. This is a case of robot behavior redesign by end users.

The first study involved cobots for manufacturing. Cobot design guidelines were developed in an iterative process, by conducting a literature search, and iteratively revising guidelines with experts in robotics and cobot User Experience (UX) design. This resulted in guidelines that designers can use to improve existing cobot UI design for EUP [99].

In a second study, an interactive prototype was developed to be able to program/animate robot motion without programming expertise. Previously, I co-developed a pose-matching imitation system for the Pepper robot [260], which was used for the purpose of robot animation in this dissertation. The goal was to develop a way to use full-body human motion demonstrations to efficiently prototype and record motions, for use in dance performances or social interaction scenarios. The prototype development resulted in interaction design considerations for systems for robot animation.

1.3.4 Transparency

The third investigation into involving end users focused on transparency. Informed by the theoretical work on interaction as described in Section 1.3.1, a user study was conducted with 31 visitors of a museum of science and technology. They were asked to interact with the Pepper robot. The robot was programmed to detect objects and communicate object locations back to the user verbally, visually, or with a combination of modalities. Participants were asked to indicate if the robot's knowledge base contained errors. The first aim was to find out how to design shared representations to support human-in-the-loop error detection. A shared representation is understood here as a shared basis of knowledge; a representation of common ground between human(s) and robot(s) involved in the interaction. The second aim was to find out how failures that occurred during the user study fit the current understanding of failure in HRI. Over the course of the interaction, some participants were found to test the system to find out more about its limitations.

1.4 Overview of Chapters

See Table 1.1 for an overview of the chapters in this dissertation with their associated research questions. In the following sections, I briefly summarize each chapter.

(Phase)	Topic	Chapter	Research question
	Theory on interaction	2	What concepts and models of communication have been applied in HRI? When arguing from the perspective that interaction between humans and robots is fundamentally asymmetric, what are design strategies we can employ that facilitate the interaction?
Design time	Co-design	3	How can we conduct participatory workshops about robotic technology components in ways that enable care home residents and care workers to envision use cases for technology in their care?
	EUP	4	What are important design factors for cobot UIs for programming?
		5	What are interaction design considerations for a system for robot animation that implements a human-humanoid imitation system?
Use time	Transparency	6	How to support human-in-the-loop error detection in an object organization task with a robot? How do the failures that occur with a functional system fit into current understanding of failure in HRI?

Table 1.1: Overview of the chapters.

1.4.1 Chapter 2: Communication Models in Human-Robot Interaction

Research Question Chapter 2. *What concepts and models of communication have been applied in HRI? When arguing from the perspective that interaction between humans and robots is fundamentally asymmetric, what are design strategies we can employ that facilitate the interaction?*

What are underlying theoretical assumptions when we talk about interaction or communication in HRI? The article identifies different ways communication and sociality with robots are conceptualized in the literature. Anthropomorphic design strategies are frequently applied to capitalize on the human tendency to anthropomorphize [34], with the idea that humanlikeness will mean that people semi-automatically know how to interact with the system because they can apply schemas from human-human interaction. In this chapter, I specifically argue for seeing interaction between humans and robots as asymmetrical. Capabilities are not symmetrically distributed. For example, there are differences between humans and robots regarding perceptual processing and action capabilities.

Based on these asymmetries, I argue that this makes it important to design the robot's embodiment and behavior in a way that makes the robot's capabilities explicit. This can be achieved by signifying how the robot can be interacted with, and communicating the robot's perceptual, action, data processing, and decision-making capabilities (transparency and explainability). I focus on collaboration processes, in which external representations of the joint activity can support the end user in keeping track of task progress (e.g., using a GUI).

1.4.2 Chapter 3: Co-design of Robotic Technology with Care Home Residents and Care Workers

Research Question Chapter 3. *How can we conduct participatory workshops about robotic technology components in ways that enable care home residents and care workers to envision use cases for technology in their care?*

For the co-design workshops, a mutual learning format was developed, where researchers gave technical input and then discussed potential applications of the technology in the care context with care home residents and care workers. This enabled the researchers to learn more about what is important for care stakeholders. We developed a series of five workshops: one where we discussed admission to the care home, one on robots, one on computer vision, one on conversational AI, and a final workshop on participatory design. The workshop series was conducted three different times with different group configurations. In total, 13 care workers and 12 care home residents took part. For the workshops, handouts, explanations, and worksheets were prepared, as well as technology demonstrations. One technology demonstration covered object detection and pose detection, while during the other one participants could ask a conversational system questions. The takeaways from the workshop series are as follows. (1) Including both care workers and residents offers complementary perspectives. (2) An interactive conversational prototype sparked most engagement from residents and facilitated joint sensemaking of the technology. (3) There are both advantages and disadvantages to leaving the application scenario open. While it leaves more space for identifying stakeholder needs, it can also be challenging to imagine applications if the scenario is open (4) Group activities offer the advantage of different perspectives, but potential group dynamics require consideration (e.g., hierarchies among care workers). Documentation Support was an application scenario that was frequently mentioned by care workers that could potentially be supported by technology.

1.4.3 Chapter 4: Design Guidelines for Collaborative Industrial Robot User Interfaces

Research Question Chapter 4. *What are important design factors for cobot UIs for programming?*

This study involved collaborative robots (cobots) that are programmed with kinaesthetic teaching (lead-through programming) [4], which involves physically moving the cobot arm while recording intermediate points to program them, or programming them with a teach pendant (a physical device with a GUI), usually in a manufacturing context. This type of interaction is often described as fast, intuitive, or straightforward and as lowering cognitive load [4, 77]. However, the GUI design of current systems leaves room for improvement [5, 243]. For this case study, existing cobot UIs were tested, literature on design factors for cobot and robot UIs was reviewed, and cobot UX experts in industry and academia were interviewed. Based on these different sources, design guidelines for cobot programming UIs were developed that can be used for heuristic evaluation [99].

1.4.4 Chapter 5: Programming Robot Animation Through Human Body Movement

Research Question Chapter 5. *What are interaction design considerations for a system for robot animation that implements a human-humanoid imitation system?*

This study involved the interaction design of a system for programming (or animating) a humanoid robot's motion (Pepper) by human demonstrations. The goal was to develop a way to be able to use full-body human motion demonstrations to easily record motions for the Pepper robot, for use in dance performances or social interaction scenarios. A prototype system for robot animation was developed on the basis of a pose-matching imitation system [260], using human demonstrations to record robot motions by means of Kinect and a GUI. Based on the development of the prototype, interaction design factors of such systems were outlined according to an interaction design framework for robot animation. For this purpose, an interaction design framework for computer animation [293] was extended to be suitable to describe the development of robot animation using human-humanoid imitation systems.

1.4.5 Chapter 6: Human-In-The-Loop Error Detection in an Object Organization Task with a Social Robot

Research Question Chapter 6. *How to support human-in-the-loop error detection in an object organization task with a robot? How do the failures that occur with a functional system fit into current understanding of failure in HRI?*

This study involved comparing different ways of communicating object positions detected by a robot, to understand how participants construe the way the robot works and to find out what their preferences are in relation to the way the robot's knowledge is communicated. In a within-subject user study (N=31), different representations of detected objects were conveyed back to the study participant. This was done in the form of speech, by means of visualization on the tablet, or through a combination of both speech and visualization (multimodal condition). These representations supported error detection in the study task. Participants preferred multimodal representations. In the study, participants increased the complexity of the way they organized objects to determine if the robot would still detect objects correctly. In other words, they appeared to test the system limitations. This trial-and-error behavior challenges existing classifications of failure in HRI that tend to focus on failures that are purely technical in nature or stem from user behavior (usually focusing on deliberate violations or mistakes). However, the type of failure we described stems from the combination of technical limitations, objects in the environment, user motivations (e.g., curiosity) and user actions. This is a more holistic understanding of failure in HRI.

1.5 Contributions

Overall, **this cumulative dissertation contributes to a better understanding of interaction design factors and methods that enhance human agency in HRI scenarios**, specifically by making it possible to impact robot behavior.

The High-Level Expert Group on Artificial Intelligence published ethics guidelines for trustworthy AI (also applicable to embodied AI) [124]. They argue that the development, deployment, and use of such systems should meet human agency, human oversight, and transparency requirements. This dissertation contributes to an improved understanding of how such ethics guidelines can be translated into practice. The report's human agency requirement states that users should be supported in decision-making with regard to AI systems [124]. Through the provision of knowledge and tools, they should be supported in understanding and interacting with AI systems. Systems should support user decision-making that aligns with user goals and allow for some form of human oversight. The transparency requirement in the AI ethics guidelines [124] states that the system's decision-making process should be traceable (traceability) and understandable to humans (explainability). Users should be informed that the system they are interacting with uses AI and should be informed regarding system capabilities and limitations. In this dissertation, these requirements come back in the following ways. In Chapter 2 I reflect on transparency and indicating system capabilities in HRI. The co-design workshops in Chapter 3 included discussions of robot strengths (e.g., precision) and weaknesses (e.g., empathy), as well as the capabilities and limitations of computer vision and conversational AI technologies. Cobot UI guidelines in Chapter 4 included, for instance, system state awareness and accessibility of information to end users. In the study in Chapter 6, representations were developed to give users information regarding objects that are currently perceived by a robot.

The contributions of the individual chapters are as follows. The contributions of Chapter 2 are (1) an overview of communication models as applied in HRI, (2) the argument to understand HRI as asymmetric, (3) an asymmetric model of HRI, and (4) design recommendations based on this model and asymmetric understanding. For the co-design study in Chapter 3, the contributions are (1) the design of the co-design workshop series and (2) recommendations for co-design workshops with care home residents and care workers in different group constellations. The focus of Chapter 4 is on cobot UIs that are used for reprogramming the cobot. The contribution of the chapter consists of design guidelines for such cobot UIs. For the study on robot animation as described in Chapter 5, the contributions are (1) an implementation of a prototype for robot animation based on a human-humanoid imitation system and (2) interaction design considerations for such systems described according to a design space for robot animation. The contributions of the transparency study in Chapter 6 are (1) multimodal representations of object configurations detected by a robot that can be used for human-in-the-loop error detection, (2) results from a user study on these representations, and (3) extension of failure understanding in HRI with the concept of productive failure.

I discuss the contributions of the dissertation across the chapters below. One component of enhancing human agency through interaction design concerns supporting learning how technology works using formats for participation and mutual learning, and designing the robot's behavior to disclose its capabilities (Section 1.5.1). Furthermore, a contribution is the consideration of how human agency can be supported by offering possibilities to affect robot behavior (Section 1.5.2).

1.5.1 Disclosing robot capabilities and supporting user learning in interaction

One cross-cutting theme was supporting user learning through interaction design. This required enabling users to build up an understanding gradually, and creating spaces for participants to test their interpretations of the way a system works within the context of interaction. Learnability can be connected to principles such as simplicity, predictability, and familiarity [297, p.3]. A familiarity argument has been made for humanlike design (e.g., [120]). There are limits to relying on already familiar interaction concepts from interaction between people. In this dissertation, the argument is developed that interactions of humans and robots are asymmetric regarding the capabilities of humans and robots involved (Chapter 2). Robot capabilities, task and behavior may not be immediately apparent to human interaction partners. To facilitate interaction and end user impacts on robot behavior, I argue in this dissertation that the behavior and design of robotic systems should disclose robot capabilities and possibilities for interaction. Metaphors from interaction with other technologies and objects can similarly be familiar and can also be combined with metaphors from interaction with people.

In this way, learning about technology in the context of interaction can contribute to the user's *technology literacy*: *"an individual's abilities to adopt, adapt, invent, and evaluate technology to positively affect his or her life, community, and environment"* [117, p.117], which is argued to promote self-efficacy [117, p.117]. Self-efficacy beliefs, or people's self-judgment regarding their capability to achieve intended actions, may be impacted through mastery experiences such as interacting with a robot or observing others do so [219]. Davies [63] proposes a framework for technology literacy. At the awareness level, the person is exposed to technology and discovers what it can do. The praxis level involves trying out the technology, finding out how it can be used, and how functionalities can be achieved. The phronesis level involves being able to reflect on why it is appropriate or not to use the technology in a given situation. In the dissertation, participants could try out technologies and discover their possibilities and limitations (as in, e.g., Chapters 3, 5, 6) and reflect on its use (as in, e.g., Chapter 3). For example, in Chapter 3, formats for mutual learning were developed to support stakeholders in learning about technologies and researchers in learning about stakeholder needs. Formats included technology presentations (on Computer Vision and Conversational AI), discussion formats, drawing activities, and direct interaction with a conversational prototype. In this study, it was important to design the information materials and workshop formats to be suitable for the target audiences. In Chapter 6, a user study is described in which errors were observed that played a role in user learning regarding the way the system worked. Users tested the system's limitations, provoking some errors in the process. Communication and education about AI technologies, as well as stakeholder participation, can play a role in establishing trustworthy AI [124].

1.5.2 Supporting end users in affecting robot behavior: Constraints on design choices and mutual effects

One aim was to enable end users and other stakeholders to affect robot behavior and make it more context-appropriate. In the dissertation, this required explicitly creating technical and social possibilities for changing robot behavior. In this section, I describe how end users and other stakeholders could impact robot behavior in different ways in the dissertation. The possibilities for participants in the studies to affect robot behavior included discussion formats (co-design study in Chapter 3), consideration of interaction design aspects that make it easier to interpret and impact robot behavior (Chapter 4, 6), and building systems for impacting robot motion (Chapter 5). In Chapter 5, the possibility to include tacit knowledge in programming robot motion was investigated. A prototype was built for human-humanoid motion imitation that allowed for recording of physical movement.

One part of the dissertation concerned the relation between system architecture, system capabilities and interaction design (especially Chapter 2, 5, and 6). The system's capabilities are determined by programming and the system architecture (input modalities, output modalities, how the system can be interacted with, how data is processed, how information is presented to the users). I considered how system architecture choices affect interaction design, and how interaction design can make it easier to interpret the system architecture, to better enable end user impacts on the system. The system architecture choices that could be made in relation to the robot animation prototype directly affect how the participant is able to interact with the system (Chapter 5). In the dissertation it is apparent that in making design choices, many stakeholders and technologies are at play that pose constraints on and mutually influence one another. For instance, one participant in the interview study on cobot systems reported cultural differences regarding allowed information access by users (Chapter 4). Such mutual effects and mutual constraints of different agents and system architecture choices highlights the complexity of the question of agency (see Section 1.6.1).

What can be observed in the application of the different methods in the dissertation is that choices by participants were increasingly constrained in going from design time to use time (see Figure 1.1). In the co-design study (Chapter 3), the application scenario was still open and participants were asked to speculate on potential use cases. Here, the space of design options is still large, and many design choices can potentially be made. At the same time, this openness was experienced as challenging by some participants. Disadvantages include that while many ideas may be generated, there is limited capacity to incorporate all ideas, and some ideas may not be feasible. The choices that can be made in a co-design process prior to determining the application of a technology allow for complete reconsideration of the problem framing, while other applied methods assume the existence of a particular technology and task. Participant considerations during co-design workshops can be translated into system requirements later on and can thereby affect system architecture choices. While the EUP environments in Chapter 4, 5 enable a direct impact on robot behavior, in the end, these only allow for changes that are possible within the confines of technical choices that have already been made, such as the programming possibilities that the EUP environment allows for. Programming environments encode and enable specific forms of (social) interaction. System architecture choices that have already been made open up and close large spaces of potential design alternatives. Conversely, while the impact in the transparency study (Chapter 6) was limited to confirming the configuration of objects, the interaction that takes place or can be expected to take place in the “use phase" is much more pronounced. This study focused on a specific scenario, namely an object organization task. The design space of possible choices that can be made in realizing design [31] differs across a design process, or in other words, different parts of the design space are under consideration.

1.6 Discussion

1.6.1 Agency

When an objective such as enabling end users to impact robot technology is achieved, the “impact" does not only go one way, from user(s) to robot(s). Robots also act on their environment. Moreover, the design of technology involves making choices that influence humans who use or otherwise relate to these technologies. Several authors discuss this as a form of mutual shaping. Šabanović [312] proposes the mutual shaping framework in which society and robotic technology mutually influence each other. Technologies have affordances that affect their use in society, while in the process of designing technology, technical and social choices are made. Verbeek [290] argues that many relations people have with technologies cannot adequately be described as use relations. Designers design not just products, but also human practices, experiences, and human-world relations. Verbeek [289] argues that technologies mediate human actions, as they have an impact on how humans interpret their reality and how they can decide and act. Technologies are thus not neutral. Verbeek outlines different forms of agency: the agency of the human interacting with the artifact, the agency of the technology designer, and the agency of the artifact itself through its mediating role regarding human action. While the technology’s action is not necessarily deliberate, it directs human action. The intentionality of the technological artifact only exists through this technological mediation of human action and decision-making, constituting hybrid intentionality. Due to the ability to physically affect and navigate the environment, and the increased capability of robotic systems to use language, this has the potential to mediate human action even further (than other technologies or objects) and affect human norms in the process [144].

Several authors have written about the question if (social) robots have (social) agency. In discussions of agency, a tension can be felt between agency, the existence of intentionality, and perceptions of agency. An agent is, in the context of AI and robotics, usually understood as *“anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors”* [237, p.31]. In this sense, robots are also agents that act on their environment. In contrast, the concept of agency often presupposes internal states such as intentions. Alač [6] describes a social robot as being both a thing and an agent, and describes these aspects as entangled. Jackson and Williams [144] aim to avoid internal aspects such as intentionality in their definition of social agency. Jackson and Williams [144] write that agency requires autonomy, interactivity, and adaptability with regard to observables at a certain level of abstraction (LoA), e.g., the user’s or the developer’s LoA. Social agency requires agency and social action. Social action affirms or threatens face of a social patient. For an action to be social, it thus requires a face-affecting action, a social agent, and a social patient [144]. According to this understanding, people can be both social agents and social patients, but it is unclear to what extent robots can be. Jackson and Williams also state that it is possible for a robot not to have face at the developer’s LoA, while having it at the user’s LoA [144]. Moreover, whether face is affected or not, is open for interpretation. *Perception of agency* may remain a more appropriate concept. It has been argued that perception of agency in HRI is becoming more important as robots are increasingly social and capable [280]. Zafari and Koeszegi, who write about machine agency as an attributed capacity [307]. Trafton et al. write: *“People perceive agency* *in another entity when the entity's actions may be assumed by an outside observer to be driven primarily by its internal thoughts and feelings and less by the external environment."* [280, p.3].

While the agency of humans and machines mutually affect one another, I argue that asymmetry remains, at this point in time. In the double dance of agency model by Rose and Jones, humans and machines exhibit agency. Human and machine agency are seen as having different properties, but also as interrelated in terms of process, outcomes, and sociomaterial conditions under which interactions between humans and machines take place [234, 307]. Suchman argues that there is *"a durable asymmetry among human and nonhuman actors."* [262, p.11]. The combination of asymmetry, perceived agency, and the potential to affect human norms mean that it is also important on a societal level to set expectations of (robotic) technology at an accurate level, communicate about the possibilities and limitations of such technologies, and communicate existing external influences.

1.6.2 Inter- and transdisciplinarity

As stated before, HRI is a multidisciplinary research field. Research can take place within disciplines, or in interdisciplinary or transdisciplinary ways. Multidisciplinarity juxtaposes disciplines without aiming to integrate their perspectives [265]. Interdisciplinarity, on the other hand, involves *"communication and collaboration across academic disciplines"* [145, p.44]. Interdisciplinarity does aim for a shared understanding, answering shared questions, and integrating methods and knowledge from different disciplines. This requires team members to develop an understanding of the perspectives of team members from other disciplines [265]. Transdisciplinarity is understood here as collaboration across multiple academic disciplines and involving non-academic stakeholders to aim to answer complex questions in a contextualized way [265]. Similar to Blackwell's argument for viewing HCI as an inter-discipline or trading zone in which researchers negotiate between and collaborate across disciplines [28], HRI can be viewed as such. For instance, social robots have been described as "boundary objects" in collaboration across disciplines, which means that social robots provide a common focus while functioning as a relevant object of research within the individual disciplines that are involved [313].

The work that was conducted for this dissertation, took place in the context of the interdisciplinary Doctoral College TrustRobots and the transdisciplinary project Caring Robots // Robotic Care. In these contexts, I worked with people with different sets of expertise, such as expertise on cobots for manufacturing, HCI, sociology, and computer vision. People with practice expertise were also included in these transdisciplinary projects. Chapter 3 involved stakeholders from the care context and Chapter 5 dancers. In short, conducting the work in the dissertation was an exercise in operationalizing inter- and transdisciplinarity. What I found important, especially in finding topics to collaborate on and in supporting project coordination, was facilitating communication, mutual learning and taking the initiative to participate in joint efforts within the project team. It also involved actively looking for shared goals and mutual benefit in collaboration with researchers and care stakeholders.

Against this backdrop, the work that was conducted for this dissertation is in line with the interdisciplinary nature of HRI research. For instance, there is a strong focus on connections between system architecture choices and interaction design. Several methods were investigated to support end users in (re)specifying and impacting robot behavior. This effort was inter- and transdisciplinary. In inter-

and transdisciplinary HRI projects, an interaction design approach is relevant and valuable, as interaction design is a point of entry for HRI that touches on system architecture choices and how the system is interpreted by, used by, and responded to by the user. It is concerned with core decisions such as a robot's task in human social space.

1.6.3 Knowledge contributions through interaction design for HRI

This dissertation involves both practical design work and theory development. In Chapter 2, a model and asymmetric understanding of interaction and communication in HRI was developed, and a translation was made to design recommendations. One such recommendation, transparency, was practically implemented in the robot behavior design in the study in Chapter 6. The robot's behavior was designed to include a "scanning motion" during which the robot moved its head up and then down in the direction of the cupboard, which indicated that the robot was performing its object detection routine. Chapter 3 includes reflections on the conducted co-design process. Design guidelines were developed in Chapter 4. In Chapter 5, a robot animation framework is outlined that was developed on the basis of developing a technical prototype. These forms of exchange between theory and practice raise questions such as: how do theoretical concepts influence system and experiment design in HRI? How can the design of individual systems contribute to theory development?

One question in design research is how individual design instances can be connected to the development of more generalized theories. This is often expressed as a tension between *"ultimate particulars"* and *"global knowledge production"*. Stolterman [261] describes outcomes of design practice to be the manifestation of 'ultimate particulars', where design instances are different in every implementation context due to differences in the organization and people involved. The concept of intermediate-level knowledge has been discussed in the HCI context. Höök and Löwgren [137] argue that design-oriented research can produce knowledge that lives someplace in-between particular instances and general theories: intermediate-level knowledge. They emphasize that individual design instances are still important in themselves, but intermediate-level knowledge can play a generative role for new design instances. As generative examples they mention guidelines and methods, while evaluative forms of intermediate-level knowledge include heuristics. They also propose quality criteria for these forms of knowledge, namely that it is contestable (novel contribution), defensible (e.g., rigorous research process), and substantive (relevant for a community and its goals) [137]. Lupetti et al. [180] argue for the concept of intermediate-level knowledge as useful towards an HRI design epistemology, so that researchers can build on findings from design research.

The concept of intermediate-level knowledge has been criticized for its perceived lack of a shared epistemological basis when considering the ends of the proposed spectrum (with 'universal' theories being associated with positivism and individual design instances with social constructivism) [97]. The in-between position of intermediate-level knowledge can further be criticized for both a potential lack of rigor as well as a loss of contextualized knowledge [97]. Moreover, a notion such as "universal" invites criticism, especially in connection to values [225].

Zimmerman et al. [310] propose a different spectrum of theory, and write that exploratory work in Research through Design can contribute to "nascent theory" that can develop to more mature theory. In HRI, an example of nascent theory is the product ecology framework by Forlizzi [94]. Redström [225, p.39] proposes an alternative, open-ended spectrum from product (what a design is) to paradigm (what designing is). The spectrum includes product - project - program - practice - paradigm. *“Thus this is to be read not as a shift from design as a thing on one end to design as an activity on the other, but rather as the span between a distinct outcome and the overall orientation of the effort that produces such outcomes"* [225, p.39]. This characterization fits design’s normative orientation. For example, Oulasvirta and Hornbæk [205] construe design as a constructive activity that concerns making choices that change the future. They focus on the idea of “counterfactual thinking", in which theory leads to thinking of possible worlds and hypothetical events resulting from design decisions. The work in this dissertation fits the characterization by Redström [225], in that I sought to connect an orientation of a design effort (e.g., acknowledging asymmetry) to its implementation in individual studies and systems.

Design artifacts such as models and guidelines serve to document as well as inform a certain attitude to design; these document and inform choices made in design practice and its normative aspects. While there is a risk that guidelines are interpreted as too prescriptive, it remains the designer’s responsibility to interpret and adapt such theoretical artifacts to the context at hand. In acknowledging this normative aspect, I also emphasize that it is important for HRI research to outline how foundational concepts (such as social interaction and communication) are interpreted. How concepts are interpreted informs HRI research, whether made explicit or not. In this dissertation, I investigate how such foundational concepts are interpreted and operationalized, and challenge their understanding in HRI research, for instance, the understanding of failure in HRI in Chapter 6, and concepts and models of communication in Chapter 2. Theories can be used to inform design choices, identifying preferable choices, and reconsidering problems by showing new design spaces (see also [205]). With my work, I hope that developing theory and challenging existing theory helps make the understanding of foundational concepts richer and more robust when applied in practice, and that it may inform future HRI design and research.

1.6.4 Ethics

At the start of the dissertation, there was no ethics committee at TU Wien, and I followed the guidelines regarding informed consent as recommended by TU Wien. The studies reported in Chapter 3 and Chapter 6 were peer reviewed by the pilot Research Ethics Committee at TU Wien. In October 2023, this committee finished its pilot phase and was officially implemented. This committee does not officially approve of the proposed studies, but provides an opportunity for reflection on the proposed study and informed consent procedure. The committee requires submission of a research ethics questionnaire containing questions regarding, e.g., human study participants and additional documents such as an informed consent form.

In the studies conducted for this dissertation, participants were informed about what data would be processed and how, that participation was voluntary and that they could opt out at any time. Participants were informed about the study and were asked to sign informed consent forms regarding study participation and data processing consent. In the cobot study, participants received a safety briefing prior to working with the cobot. For the study with the Pepper robot at the museum, the study facilitators explained the way the object detection functioned on the robot after participants interacted with the robot, and answered participant questions. The ethical aspects of conducting studies in the care context have repeatedly been discussed with the ethics committee and within the project team. Further details are included in the respective chapters.

1.6.5 Limitations

Transparency might be counterproductive for some applications. Making data collected by systems inspectable can have security or privacy implications [273]. Similarly, allowing for changing the robot's programming by end users can also have safety implications. When enabling such changes, some training to handle the technology safely will be necessary. When a robot functions in a (semi-)public context, the type of changes that end users can make will be limited.

One aspect of the dissertation was the rather strong focus on cognitive aspects, namely interpreting and learning about technology. Some people may not be interested in how technologies function or in programming. Learning about technologies or programming them should not be imposed on people. Moreover, finding ways to measure how people interpret a technology deserves more study. Robotic technology is not (financially) accessible to everyone. The question remains if the development of better programming tools is enough to allow for participation in end user programming practices due to the cost and accessibility of robotic platforms.

Another limitation is that I mainly focused on dyadic interactions. Increasingly, there is attention in the HRI community for non-dyadic interaction [244]. Van Wynsberghe and Li [287] explicitly propose to reframe the model of dyadic interaction in HRI to one of human-robot-system interaction. Their model construes bots as mediators between care recipients and the healthcare system. Vallès-Peris and Domènech [284] develop the approach "Caring in the In-Between", in which they construe robots as embedded in a network. A recent article proposes the research area of Interaction-Shaping Robots, which is concerned with studying how robots influence the interaction between multiple other (human and/or robotic) agents [104]. In the co-design study in Chapter 3, non-dyadic interaction was explored to some extent, with a conversational prototype being part of a focus group. Here, the system was an object of the group discussion that participants also interacted with. The technology thus facilitated interaction between people in this group setting. However, the main type of interaction under consideration in this dissertation was dyadic interaction, with emphasis that robot interaction partners are open to outside influence (Chapter 2).

Instead of having one use case under study and investigating it in different phases of the design time - use time continuum, this dissertation explored a variety of contexts and robotic platforms. This enabled study of different methods while keeping the focus on end-user impacts on robot behavior. For future work, it could be interesting to study a specific robot application from design time to use time.

The COVID-19 pandemic limited end user inclusion, which led to choices to focus more on theory development in Chapter 2, and limited (online and in-person) involvement of domain experts in Chapter 4 and 5. Once it was possible again, studies were conducted for Chapter 3 and 6. More inclusion of end users and more evaluation of developed prototypes could however have been beneficial, especially for Chapters 4 and 5.

1.7 Publications that are part of the dissertation

See Appendix A for the clarification of the contributions to the publications that are part of the dissertation. Supporting materials can be made available upon request.

Frijns, H.A., Schürer, O., Koeszegi, S.T. (2021) *Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI)*. Int J of Soc Robotics 15, Issue date: March 2023, pp. 473–500.
https://doi.org/10.1007/s12369-021-00785-7 (Chapter 2)

Frijns, H.A., Vetter, R., Hirschmanner, M., Grabler, R., Vogel, L., Koeszegi, S.T. (2024) *Co-design of Robotic Technology with Care Home Residents and Care Workers*. To appear in: Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA'24), ACM, 10p. (Chapter 3)

Frijns, H.A., Schmidbauer, C. (2021) *Design Guidelines for Collaborative Industrial Robot User Interfaces*. In: Ardito, C., et al. Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12934. Springer, Cham, pp. 407–427.
https://doi.org/10.1007/978-3-030-85613-7_28 (Chapter 4)

Frijns, H.A., Stoeva, D., Gelautz, M., Schürer, O. (2024) *Programming Robot Animation Through Human Body Movement*. To appear in: Proceedings of the 2024 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO), 7p. (Chapter 5)

Frijns, H.A., Hirschmanner, M., Sienkiewicz, B., Hönig, P., Indurkhyam B. and Vincze, M. (2024) *Human-In-The-Loop Error Detection in an Object Organization Task with a Social Robot*. Frontiers in Robotics and AI, Sec. Human-Robot Interaction, Volume 11, 17p.
https://doi.org/10.3389/frobt.2024.1356827 (Chapter 6)

1.8 Supporting publications

Frijns, H. A., Schürer, O. (2022) *Design as a Practice in Human-Robot Interaction Research*. Book chapter. In S. T. Köszegi, M. Vincze (Eds.), Trust in Robots, TU Wien Academic Press, pp. 3–29.
https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_1

Stoeva, D., **Frijns, H.A.**, Gelautz, M., Schürer, O. (2021) *Analytical Solution of Pepper's Inverse Kinematics for a Pose Matching Imitation System*, RO-MAN 2021, 8p.
https://doi.org/10.1109/RO-MAN50785.2021.9515480

Dobrosovestnova, A. and **Frijns, H. A.**, Brauneis, C., Grabler, R., Hirschmanner, M., Stoeva, D., Vetter, R., Vogel, L. (2022) *Transdisciplinary and Participatory Research for Robotic Care Technology - Mapping Challenges and Perspectives*, Workshop CROA (Care Robots for Older Adults), RO-MAN 2022, Naples, Italy, 3p. https://doi.org/10.34726/2862

Frijns, H.A., Schürer, O. (2020) *Context-Awareness for Social Robots*, IOS Press, Frontiers in Artificial Intelligence and Applications, Vol. 335: Culturally Sustainable Social Robotics, pp. 520 - 524.
https://doi.org/10.3233/FAIA200951

Frijns, H.A., Schürer, O. (2020) *The impact of human-like and device-like signals on people's mental models of robots*, Workshop Mental Models of Robots at HRI 2020

Hirschmanner, M., Grabler, R., **Frijns, H.A.**, Mayer-Haas, E., Vincze, M. (2024) *Prototype of a Care Documentation Support System Using Audio Recordings of Care Actions and Large Language Models*, Workshop on Human-Large Language Model Interaction, HRI '24, March 11–15, 2024, 3p.

Grabler, R., Hirschmanner, M., **Frijns, H.A.**, Koeszegi, S.T. (2024) *Privacy Agents: Utilizing Large Language Models to Safeguard Contextual Integrity in Elderly Care*, Privacy-Aware Robotics Workshop, HRI '24, March 11–15, 2024, Boulder, CO, US, 5p. https://doi.org/10.34726/5960

CHAPTER 2

Communication models in Human-Robot Interaction: An Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI)

This dissertation chapter has previously been published as the following publication: Frijns, H.A., Schürer, O., Koeszegi, S.T., Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI), International Journal of Social Robotics 15, 473–500 (Published May 3 2021, Issue Date March 2023), Springer.
https://doi.org/10.1007/s12369-021-00785-7

2.1 Abstract

We argue for an interdisciplinary approach that connects existing models and theories in Human-Robot Interaction (HRI) to traditions in communication theory. In this article, we review existing models of interpersonal communication and interaction models that have been applied and developed in the contexts of HRI and social robotics. We argue that often, symmetric models are proposed in which the human and robot agents are depicted as having similar ways of functioning (similar capabilities, components, processes). However, we argue that models of human-robot interaction or communication should be asymmetric instead. We propose an asymmetric interaction model called AMODAL-HRI (an Asymmetric MODel of ALterity in Human-Robot Interaction). This model is based on theory on joint action, common robot architectures and cognitive architectures, and Kincaid's

model of communication. On the basis of this model, we discuss key differences between humans and robots that influence human expectations regarding interacting with robots, and identify design implications.

2.2 Introduction

In the Human-Robot Interaction (HRI) literature, models of interpersonal communication have been applied (see for instance [167][300]), and several models for human-robot interaction and communication have been developed (e.g. [122][149][64]). Concepts from communication theory have been discussed and applied, but often with little theoretical context. The present paper aims to fill this gap through thorough reflection on existing models in the literature on communication between humans and HRI, in order to (1) connect existing models and theories in the field of Human-Robot Interaction to different traditions within communication theory, (2) critically discuss (symmetric) models of HRI and (3) formalize an asymmetric model for human-robot joint action. Our main aim is to make the asymmetries between a human and a robot agent that are engaged in a communication process explicit, in order to provide design guidelines that mitigate potential communication failures arising from these asymmetries.

The first aim of this article is to connect communication theory with HRI and social robotics, an interdisciplinary endeavour. Researchers in HRI already apply concepts from semiotics [121], for instance, but the connection to a broader research field and theory of what constitutes communication is often lacking. We aim to go beyond simply borrowing concepts and models from communication research, and instead connect these concepts and theories to the broader field of communication theory. This will hopefully give HRI researchers a better overview of entry points and existing theory on models of human communication. We also wish to point to the potential of communication theory as a practical discipline that can serve to inform the design of robotic systems.

The second aim is to critically discuss communication models as currently applied in HRI. We identify shortcomings of existing models. We posit that current communication models of HRI are lacking with respect to the context dependency of interactions, the influence of external actors, and asymmetry between humans and robots. With asymmetry, we mean that robots and humans are at present fundamentally different entities, and rather than focusing solely on their (theoretical) similarities in models and designs, we should carefully consider their differences. Models for communication between two agents that are 'symmetrical' (reflectional or bilateral symmetry) presuppose that we can use the exact same components to model any agent and that both agents have similar requirements and ways of functioning within the interaction. An asymmetric model, on the other hand, does not assume this. We argue that acknowledging differences between humans and robots should be embedded in models and robot designs. Other scholars have similarly argued that human-robot interaction should be conceived as asymmetric [161][247], and emphasize that there are functional, physical and cognitive differences between humans and robots [58]. Guzman and Lewis [113] argue that the similarities and differences between humans and robots need to be assessed. When it comes to modelling human-robot communication, other researchers have argued that a new model of human-robot communication is required: a model including facets of communication that remain implicit in existing models (for example, knowledge of mission goals and cultural norms [300]). While the concept of asymmetry has previously been proposed by other researchers [119][161][240][247], the consequences of this concept have not been analysed in detail with regards to communication modelling and interaction design, which is what we set out to do here. See Section 2.3.3 for a more detailed discussion of the concept of asymmetry.

Building upon this second aim, we propose a model for human-robot joint action and communication, our third aim. We use this model to discuss the differences between the human and the robot side of the model, and highlight design implications based on the model and the identified differences. We will highlight how this model is asymmetric and how both the human and the robot agent can be influenced by external actors. We argue that such a model contributes to an enhanced understanding of key differences between humans and robots, and how these differences can conflict with human expectations of the interaction. This, in turn, can help us reconsider robot design (behaviour and embodiment) and increase usability. The model is called AMODAL-HRI (Asymmetric MODel of ALterity in Human-Robot Interaction). The name ‘AMODAL’ is deemed fitting, as amodal completion (or *“[t]he perception of complete objects behind occluders"* [78, p.1188]) refers to the phenomenon of perceiving an object as whole even if it is only partially perceivable. A similar phenomenon can occur with respect to robots: people may perceive a robot as a being with agency, even though it is only made up of a collection of technical components. By introducing a “model of alterity”, we uncover and dismantle the differences between human and robot actors, which allows for making the best of use of their complementary capabilities. *Alterity*, or otherness, refers to the term alterity as developed in phenomenology. The phenomenologist Don Ihde uses the term *alterity* to describe a particular set of human-technology relations, specific to relating to *“technology-as-other”*. The *quasi-otherness* that Ihde describes, indicates that some technological artefacts occupy a status between objects and human or animal otherness [138]. This understanding of certain technologies as *quasi-other* leads to the understanding of relations to those technologies as alterity relations [139]. We emphasize that this is the sense of ‘otherness’ referred to in the model name, not a human otherness. Otherness applied to the technological artefact refers to the fundamental otherness of the robot as a sociotechnical assemblage that is subject to outside control. See also Section 2.3.3.

We argue that connecting communication theory on human-human interaction to HRI and highlighting asymmetries is important, firstly because a robot is an embodied entity that acts (to some extent) autonomously in physical space, with actions that are communicative to a human interaction partner (and those actions can be expressive of, or be interpreted to be indicative of, agency), and secondly because of the anthropomorphic design strategies that are employed in social robotics and HRI. We argue that there are both advantages and disadvantages to modelling robot communicative behaviours based on human ones. The advantages are that theories on communication and interaction between humans give us significant insight into what humans might expect from a robotic interaction partner and what is necessary for their collaboration to be successful. Krämer et al. [163] argue that there is no real alternative to using theory from human-human interaction, as humans will expect communicative mechanisms similar to those they are familiar with from interactions with other humans, though we should ensure that the theory we use is applicable to a device or robot context. Potential disadvantages are (1) that human expectations of a robotic interaction partner may be too high if humanlike communication behaviours are implemented on the robot [41], and (2) that a focus on humanlike embodiments and behaviours may be restrictive (using alternative design strategies may lead to more diverse, more successful designs) [240]. Therefore, we argue that acknowledging differences between humans and robots rather than pursuing similarity alone is a relevant additional design strategy in order to (1) avoid communication failures when a human interacts with a robot and expects human-level functioning, and (2) expand the range of possibilities and make it explicit that we do not necessarily need to model robot bodies and behaviours on human ones.

The structure of this article is as follows. First, in Section 2.3, we discuss definitions of communication and interaction as well as key research traditions in communication theory. We will also explain what we mean with asymmetry in more detail. In Section 2.4, we discuss general communication models that have been developed to model interpersonal communication and their application to HRI. In Section 2.5, we discuss models that have been specifically developed in the context of HRI and the different types of interaction they represent. In Section 2.6, we discuss different ways interaction and communication are conceptualized in HRI. In Section 2.7, we propose an asymmetric model for HRI (AMODAL-HRI) that is intended to be used for comparing human and robot agents, and identifying differences in capabilities that can lead to problems regarding human expectations of the interaction. We identify differences between the human and the robot agent in our model and give several design recommendations based on those differences in Section 2.8.

Note that our discussion will focus on models of communication and their application to HRI and social robotics. Additional topics, such as sociality, language, signs and signals are discussed as important aspects within models of communication that have received attention in the context of robotics, but are not the main focus of this article.

2.3 Interaction and communication: What do these terms mean in the context of HRI?

In this section, we first provide some theoretical background from the field of communication theory. In Section 2.3.2, we discuss the definition of interaction between one or more humans and one or more robots. We also discuss whether the term ‘communication’ applies to interactions between humans and robots. In Section 2.3.3, we explain the concept of asymmetry in HRI.

2.3.1 What is ‘communication’?

In this section, we establish that there are several different traditions in communication theory, and that each tradition views communication in different ways. Definitions of communication differ in their level of abstraction, whether they describe communication as intentional (having a particular goal) and whether the definition includes a normative evaluation (for instance, effectiveness) [177]. A multitude of definitions are possible, depending on the goal of the person who proposes the definition. When we talk about communication, we can discuss this topic at different levels of detail. Littlejohn and Foss distinguish between the level of the communicator, the message, the conversation, the relationship, groups and organizations, the media, and finally society and culture, at increasing orders of magnitude and complexity [177]. Guzman and Lewis [113], whose work is situated in the research area of Human-Machine Communication, argue that interaction with AI departs from traditional communication theory, as AI technologies have begun to take on the role of the communicator, a role that, in communication theory, could previously only be performed by humans. We discuss communication at different levels throughout this paper, but focus on the levels of the communicator, the message and the conversation.

Craig proposed the constitutive theory or constitutive metamodel of communication [57]. This model does not directly describe the communication process (it is not a first-order model), but is rather a second-order model that incorporates different views and traditions in communication research. Craig proposed that several traditions can be distinguished within the broad research field of communication theory. Communication had been an object of study within many domains, but no clear discipline had emerged by the time Craig wrote his article. The constitutive metamodel can be understood as an attempt to frame communication theory as useful as a metadiscourse (discourse regarding discourse) and as a practical discipline, oriented to the discussion of practical, real-world phenomena [57]. The traditions that Craig distinguishes are the rhetorical, cybernetic, semiotic, phenomenological, sociopsychological, sociocultural, and critical traditions in communication theory [177]. These traditions are not to be seen as incompatible or completely separate; combinations and overlaps are common. In each tradition, communication is understood in a different way:

- Rhetorical tradition: communication as a *“practical art of discourse"* [57, p.135]
- Semiotic tradition: communication as *“intersubjective mediation by signs"* [57, p.136]
- Phenomenological tradition: communication as *“dialogue"* or the *“experience of otherness"* [57, p.138]
- Cybernetic tradition: communication as *“information processing"* [57, p.140]
- Sociopsychological tradition: communication as *“a process of expression, interaction, and influence"* [57, p.143]
- Sociocultural tradition: communication as *“a symbolic process that produces and reproduces shared sociocultural patterns"* [57, p.144]
- Critical tradition: communication as *“discursive reflection"* [57, p.147]

Some of these traditions may be more useful to HRI researchers than others; however, it should be kept in mind that one tradition by itself will not suffice for describing all that is relevant to communication. With regard to HRI, we can distinguish lines of research applying views of communication that can be linked to the traditions identified by Craig. The views that are common in HRI are mostly in line with the semiotic, cybernetic, sociopsychological and sociocultural traditions[1].

[1] For a discussion of how semiotics is studied in the context of HRI, see Section 2.6.1. Theories that fall within the sociopsychological tradition of communication theory, such as communication accommodation theory and interaction adaptation theory [177], have been studied in the context of HRI as well, see interactional synchrony for example [179]. For work in HRI in the sociocultural vein, see Section 2.5.2. Sandry applied traditions in communication theory to different types of robots in order to analyse them. For instance, she analyses the robot Kismet using the sociopsychological and sociocultural traditions [240].

In the semiotic tradition, the capability to use (human) language[2] is of particularly high importance. However, we can also speak of communication in other animal species [176], and animals such as dolphins have been shown to be capable of at least some aspects of language comprehension and production [206]. Besides the capacity to use verbal or written language or other symbol systems, we consider nonverbal behaviour, body language, and other ways of signalling to be part of communicating. Nonverbal behaviour includes gestures and body movements (including, for instance, eye gaze), proxemics, touch, and appearance [275].

2.3.2 Can we speak of 'communication' with respect to interactions between humans and robots?

Before discussing human-robot communication specifically, how might we define *interactions* between humans and robots? In the context of HRI, Bensch et al. propose a model that treats interaction as *"an interplay between human(s), robot(s), and environment"* [25, p.184]. Goodrich and Schultz describe interaction in the context of HRI as *"the process of working together to accomplish a goal"* [109, p.217]. They propose the concept of dynamic interaction as a characterization that incorporates all five dimensions HRI designers can affect, namely autonomy, how information is exchanged, team structure, learning and training of the humans and robots involved, and the shape of the task [109].

How does communication relate to interaction?

Goodrich and Schultz [109] take communication as a requirement for human-robot interaction; in their view, interaction is, in some form, present in every robot application. They separate both communication and interaction in HRI into the categories proximate and remote interaction. Based on their text, we infer that their view of communication is that of exchanging information. Their description of information exchange in HRI focuses on the media used in communication processes and the format of communication. Visual, auditory and tactile modalities are most relevant in HRI, and these are typically present in forms such as: visual displays, gestures, natural language (text and spoken language), audio, physical interaction, and haptics [109].

These interaction modalities can also be combined in multimodal interfaces, enabling the user to interact with the system by means of multiple communication modes, and providing the user with multimedia output. Advantages of multimodal interaction are that they can better support users' preferences, enhance the expressive power of the user interface, reduce user errors, and lead to small efficiency improvements [73]. Other reasons to use multimodal interfaces can include reducing the human's cognitive workload and increasing the ease of learning to use the interface [109].

[2] Language has been described as "a multifaceted and complex ability that allows us to assign arbitrary symbols meaning and to use and understand these symbols in referential exchanges with others that draw joint attention to agents, objects, and events both present and displaced in space and time" [206, p.1]. Lindesmith et al. describe that language has the following properties: language consists of combinations of meaningless sounds into words that are meaningful (duality) and new (productivity). The relation between symbol and signified is arbitrary (arbitrariness). The speaker has the possibility to discuss things that are not in the same time and location as the speaker (displacement), and capacity for cultural transmission (learning instead of genetic transmission of information). Words have distinct functions (specialization) and speakers can recreate and reproduce messages (interchangeability) [176].

The user interface is what allows the human to interact with the robot, or allows the human to interact with an environment using the robot. The concept of the user interface (as a restricted area reserved for information exchange) is problematized in robotics for co-located robots, as in this case we no longer have a restricted area that is reserved for interaction (such as a screen). Instead, the entire embodiment of the robot, which acts (semi-)autonomously in its environment, becomes communicative or informative to the user. In remote interaction, as found in teleoperation applications, by contrast, the user can access the robot and its environment by means of a screen and control modalities, which is more similar to traditional device operation. Here, the user interface is restricted to devices that allow the user to exchange information with the robot and to affect both the robot and its environment.

Should we use the word “communication" to describe the interaction between human(s) and robot(s)? According to Seibt, using intentionalist vocabulary to describe robots is confusing, inaccurate and imprecise. She proposes OASIS, the Ontology of Asymmetric Social Interactions, which provides a description language for further theory developments, and offers the possibility to include non-humans as social interaction partners while emphasizing differences with human social interaction. Seibt argues that in order to “work with" robots, for instance, robots would need to be capable of having the phenomenological experience of working [247]. Instead of arguing that a robot is capable of communication, we can apply Seibt's framework and say that the robot is either functionally replicating, imitating, mimicking, displaying or approximating the process of communicating. In the context of this article, however, we are mostly concerned with the human-robot system and the way the robot appears to the human (as opposed to what the robot is capable of by itself). We would argue that to a human interaction partner, the robot will come across as communicating, at the very least in the sense of exchanging information. Although it may be imprecise to speak of “communicating", we maintain the term for sake of practicality. However, we urge the reader to keep Seibt's proposal in mind. While it can be argued that there is no communication with the robot but instead with its designers or developers, this is not a sufficient explanation for communication or interaction with (semi-)autonomous systems. In such a case, there is an element of unpredictability in the interaction and how the human and the robot relate to the current situation. In such a case, the human could be said to be communicating with a dynamic system consisting of a robot and its designers, developers, maintainers, et cetera, with the communication process focused on or enabled by the embodied robot. For our purposes, we define communication in the context of HRI as actions performed by human and robotic agents that have aims such as coordinating behaviour, reducing uncertainty, and building a common understanding.

2.3.3 Asymmetry

Coeckelbergh [52] posits that human-robot relations can be described as social relations, since robots perform roles in society and participate in interactions with humans that can be described as quasi-social. He describes robots with the term *quasi-others* to indicate their appearance as social actors. Alaç [6] proposes that in certain settings, both thing-like and agent-like characteristics of social robots are present. These different characteristics surface at different moments in the interaction.

While humans can experience a robot as a social entity, this does not take away from the fact that humans and robots are very different, and that robots are at present very limited in their abilities. People may *experience* robots as social entities. However, robots do not have the same capabilities or responses as humans with respect to social interaction. Therefore, we should in a theoretical discussion highlight these differences and study how they can become relevant for the design of robots and robot behaviours. This can also help identify when expectations on the human side may arise regarding social responses by the robot.

Generally, it can be said that any two agents (humans or otherwise) who engage in a communication process are different from each other to some extent. They bring different sets of background knowledge to the interaction, different needs, different bodies, different (cognitive) abilities, etc. We can describe this as a kind of 'hidden' asymmetry that needs to be made (more) explicit when discussing communication and interaction at a high level of abstraction. However, in human-human interaction, it can also generally be assumed that there is a significant level of similarity between the interaction partners. We can assume some level of shared background knowledge when interacting with another person, we can expect that our interaction partner will abide by similar conventions, will often speak a language we are familiar with, and if not, at least have similar needs such as the need for food and water, et cetera.

The situation is different for human-robot interaction, however. In this context, we cannot assume similar processing mechanisms, similar background knowledge, or similar functional, cognitive or physical capabilities. Instead, human-robot interaction and communication are better described as asymmetric.

Other authors have made similar arguments, most notably with regards to agency (which refers to the capability to act in an autonomous way [308]). Seibt [247] argues that interactions between humans and robots *"form a new type of social interaction("asymmetric social interactions") where the capacities for normative agency are not symmetrically distributed amongst interaction partners and which therefore are not by default potentially reciprocal, as this is the case with social interactions among humans"* [247, p.135]. Kruijff [161, p.154] has argued that robots are functionally asymmetric to humans and has proposed the concept of asymmetric agency. This concept refers to a group of agents in which individual members of the group have different understandings of reality ("asymmetry in understanding"), which in turn can result in different expectations regarding ways of acting in this (differently-understood) reality. In addition to asymmetry in understanding and capacities to act, symmetry and asymmetry between humans and robots can also be identified on the level of embodiment. Sandry [240] argues that the development of humanoid robots and of human-like communication mechanisms for robots point to a pursuit of commonality and a view of communication as the transmission of information. This could be described as using 'sameness' as a design strategy, which she is critical of. Instead of striving for "complete comprehension", she argues for striving for a "partial understanding" that recognizes the *"alterity of the machine"* [240, p.8]. Hassenzahl et al. note that humans perceive robotic and AI systems as counterparts instead of tools and write *"it creates a fundamental shift from an* embodied *relationship with technology to one of* alterity: *Technology becomes other"* [119, p.54]. They call such systems *otherware*. Mimicking human or animal communication strategies comes at the risk of stereotypical designs and disappointment. They propose *animistic design* as an alternative design strategy, and point to a need for the HCI community to develop new models, interaction paradigms, design patterns and design methods for otherware [119].

Gunkel notes that *"Communication studies (...) must (...) reorient its theoretical framework so as to be able to accommodate and respond to situations where the* other *in communicative exchange is no longer exclusively human"*[112, p.2].

Sandry [240] argues that humans and robots are different entities and that this otherness of robots can be valuable, instead of a problem to be overcome. She argues that this is made difficult by the fact that in communication theory, communication is often framed as accurate information exchange or a means to reproduce shared social patterns. In other words, communication is often viewed as a means to emphasize what communicators have in common, and increase the similarity between those who are communicating. However, this comes at a loss, as the *other* has different points of view that are devalued in a communication process which is aimed at enhancing similarity. Based on work by Pinchevski, she notes that such an understanding of communication can even be described as *"violent to the other"* [240, p.5]. While this is a far greater issue in human-human communication, the disadvantage of not acknowledging differences is a potential loss of possibilities [240]. In addition, a culture of emphasizing what humans and robots have in common comes at the risk of rendering humans as computational. For HRI, this could result in missed opportunities to design other behaviours and embodiments that are not based on the human model. In human-robot teams, not acknowledging differences poses a risk for team functioning; we should acknowledge that capacities are different across the members of a human-robot team and strive to make the most of complementary capabilities. Johnson et al. propose a method of interdependence analysis to analyse the capacities of different team members and how they depend on one another, in order to make use of complementary capabilities [149]. Sandry argues that it is the difference, the complementarity of humans' and robots' skills and the coordination of these skill sets that makes collaboration successful. While humans take responsibility and usually instructs other team members, robots have other important roles to play [240].

To summarize, we view human-robot interactions as asymmetric, as the interaction partners function in different ways due to differences in embodiment, cognitive capabilities, functional capabilities, and capacities for social interaction. We expect this asymmetry to remain in place in the foreseeable future, even with large improvements to robot (cognitive) functioning. Like Sandry, we view asymmetry between human and robot interaction partners as a potentially productive, useful feature. However, we argue if this asymmetry is not acknowledged (for instance, by striving to make humans and robots function as similarly as possible, or by ignoring the existence of asymmetry), this can be problematic and result in communication failures. Assuming symmetry where there is none can be productive for initial engineering attempts, but fail to identify problems in interactions with humans. Asymmetric models are more suitable tools that can foresee at least some of these problems. Therefore, we propose an asymmetric model of Human-Robot social interactions in Section 2.7 and design recommendations based on the identified asymmetries in Section 2.8.

2.4 Classical models of communication and interaction

In this section we discuss existing models of communication between humans. In Section 2.3.1, we already introduced Craig's constitutive metamodel [57]. One can distinguish several different types of communication models, with two well-known types being transmission and transactional models.

Authors have discussed communication in different ways, depending on their goals. For HRI, a key challenge is to establish shared awareness of a team task and to coordinate actions to achieve the task goals, which is why transactional models remain relevant in this context. Other types of models that take a different perspective on communication (e.g. with a focus on power relations) can also be insightful.

2.4.1 The transmission model of communication

In transmission models of communication (or also: linear or container models of communication), communication is described as the one-way transmission of a message from a source to a receiver. These types of models serve to depict the way technological communication functions, and are used to study the process of making sure a signal arrives at its destination intact so that the original message can be reconstructed [248]. In such models, feedback and context are not considered. The message is viewed as a kind of container for meaning that is transferred from A to B (thus following a postal metaphor [42]). The most well-known transmission model of communication is outlined in the article *The Mathematical Theory of Communication* by Shannon and further developed by Weaver. Their focus was on communication systems such as telegraph, radio, and telephone systems. The model consists of a chain in which information moves from the information source, to the transmitter (which sends a signal over a channel), to the receiver, which reconstructs the message and sends it to the final destination. The message can be corrupted by a source of noise [248]. This model falls within the cybernetic tradition in communication theory.

Transmission models have been criticized for being linear and one-way [154]. Such models exhibit epistemological biases in that they treat information like a physical substance that can be carried from point A to point B, and treat minds as disembodied entities, stripped of their context. Additionally, Kincaid argues that such models focus on communication as a means of persuasion and focus on individual psychological effects rather than effects on the social whole and social relationships. A one-way model implies one-way causation; there is no space for mutual causation [154]. The signal may be corrupted by noise, but otherwise the signal should stay the same until it is decoded by the receiver. The model is useful for its purpose, which is to describe the technological process of sending a message from A to B, but not sufficient for modelling how shared meaning arises in interpersonal communication[3]. Even though the transmission model of communication has been criticized, it has frequently reappeared in the HRI literature [122][191], often with feedback loops added to turn it into a transactional model of communication.

2.4.2 Transactional models of communication

Transactional models of communication introduce the possibility for feedback from the receiver to the sender; they depict humans involved in communication as both senders and receivers. Additionally, such models often include contextual factors that influence communication. With respect

[3] It should be noted that this was not the aim of Shannon and Weaver. They were mostly concerned with the technical problem of transmitting symbols accurately across a communication system, which Weaver notes has effects on other aspects of communication, such as semantics and the effectiveness with which meaning is conveyed.

to these models, communication can be described as having the goal of arriving at mutual understanding [154] or building shared meaning and reducing uncertainty [21]. Such models often aim to identify relevant components or factors that influence human communication rather than describing the technical process of communication.

Barnlund [21] describes communication as dynamic, continuous, circular, unrepeatable, irreversible, and complex. People involved in communication *construct* meaning on the basis of the other person's messages, rather than reconstruct it. This construction of meaning should assist in deciding on a course of action that is likely to be effective and fits the demands of the current situation. Barnlund proposes 'pilot models' of a transactional model of communication. Barnlund discusses that there can be limitless cues involved in a transactional communication process (public, private, natural, artificial, behavioural verbal and behavioural nonverbal cues). Kincaid [154] applies perspectives from cybernetics to propose the convergence model of communication. This model views communication as a process. The aim of arriving at mutual understanding is achieved by creating and sharing information, which the participants in the communication process interpret. While they may converge on meaning and therefore increase their mutual understanding, they can never converge completely because each individual brings their own set of experiences to the communication process. Communication occurs within humans' psychological, physical and social realities, and building mutual understanding is supported by the actions and beliefs of both parties [154]. Because this model describes the goal of communication as reaching mutual understanding, it can also be understood as a transactional model. We will use this model later as the foundation for a model of human-robot interactions from a joint action perspective.

Classical transactional models of communications have also been discussed in the context of HRI, including the circular model of communication as proposed by Osgood and Schramm (which depicts agents as processing a message via a decoder, interpreter and encoder, and the messages between them being sent along a continuous loop, depicted by arrows between the entities) and Berlo's model of communication (with the main components source, message, channel, and receiver) [300]. Lackey et al. [167] discuss the transactional model by Barnlund in the context of HRI.

Pickering and Garrod [214] criticize transactional models of language processing that explicitly separate production and comprehension processes. Instead, they propose that production and comprehension processes need to be tightly interwoven to support agents in coordinating their actions, resulting in joint action. Agents not only predict their own actions, they also predict the actions of the agent they are interacting with. Thus, the actions of both agents can become tightly coupled [214].

2.5 Communication and interaction models in HRI

As discussed in the previous section, models describing interpersonal communication have also been applied to HRI. In this section, we review models that have been developed specifically for HRI. Some of these models are based on models from the previous section (e.g. [122][191]). One of the models we discuss in this section was developed to describe how agents can support each other's understanding through communicative actions [122], while another offers a long-term perspective on trust development and calibration in human-robot teams [64]. Models have been developed to aid in interaction design for HRI, for instance design for interdependence [149], to establish a theoretical design framework for how products evoke social behaviour [94] and to identify how different factors influence the interaction experience [306]. We distinguish different ways of describing interaction between humans and robots in the literature proposing models of HRI, namely control relationships and social interaction including collaboration. Goodrich and Schultz identify a spectrum of different levels of autonomy of the robot relative to the human in HRI, from direct control (e.g. teleoperation) to dynamic autonomy (e.g. peer-to-peer collaboration) [109]. The models in Section 2.5.1 lie on the direct control end of this spectrum, while the models in Section 2.5.2 lie on the opposite end. The models in Section 2.5.2 can be placed across the spectrum. In Section 2.5.3, we critically discuss the models and their shortcomings. As our aim is to focus on concepts of asymmetry and 'otherness' in human-robot interactions, the models that describe social interactions in Section 2.5.2 are most relevant to the purpose of this paper.

2.5.1 HRI as control

In the models in this section, the relationship between humans and robots is one of control. The human interacting with the robot determines the robot's actions. Some models in this category aim to model teleoperation, but others see general human-robot interactions as control relations.

An example of the control paradigm can be found in Yanco and Drury [305], who propose a taxonomy intended for general human-robot interactions, from situations such as controlling an unmanned aerial vehicle (UAV) to social robots. The closest thing to a model in their taxonomy are their illustrations of various configurations of a human-robot team [305, p.2843]. The interaction is understood as one involving control of one or more robots by one or more human operators. The human operators have different levels of interaction and shared agreement between them, while the robots prioritize, deconflict and coordinate tasks issued by the human controllers. They write: *"In human-robot collaborative systems, communication mode is analogous to the type or means of control from the human(s) to the robot(s) and the type of sensor data transmitted (or available to be transmitted) from the robot(s) to the human(s)"* [305, p.2841]. Sheridan [92] proposed modes of teleoperation control: direct control, supervisory control, and full automatic control. The direct control model allows the human to steer the robot and presents the human with the robot's sensor input through a UI. Supervisory control allows the human to formulate subtasks and monitor execution. The fully autonomous control model allows the human to formulate high-level goals. Mirnig et al. [191] propose a communication structure for human-robot itinerary requests in public places based on Shannon and Weaver's model, which we classify as control because the system's purpose is to respond to itinerary requests, which represents a device-like control relationship.

2.5.2 HRI as social interaction

In this section, we discuss both models of human-robot 'ecologies' and models of human-robot collaboration. We see collaboration as a specific form of social interaction, targeted at achieving a joint goal. While the models in the previous section represented an interaction similar to device operation, the inclusion of factors such as social context, joint goals, and autonomous and anticipatory action means that we need to consider aspects of experienced agency or alterity and thus asymmetry in the interaction. We discuss different conceptions of interaction in more detail in Section 2.6.

Social interaction encompasses social, emotional, and cognitive elements [109]. Dautenhahn [62] distinguished five different ways in which robots can be defined as social, namely socially evocative robots, socially situated robots, sociable robots, socially intelligent robots and socially interactive robots. Socially evocative robots rely on human tendencies to anthropomorphize, which is in line with the media equation proposed by Reeves and Nass [226]. Socially interactive robots, on the other hand, have high-level capabilities that enable them to collaborate with humans [4].

HRI as interaction in a social context

Several models of human-robot communication and interaction have been developed with the aim of supporting interaction design. Several frameworks have been developed in which the robot can perform multiple roles; the robot may function both as a social agent and as a tool. The rationale for these models is that the social context is of high relevance to interactions between humans and robots. The social context, in these models, refers to relationships between the robot and other agents and the activities they undertake in a social environment. In the Domestic Robot Ecology framework [264], for instance, the social context can be taken to refer to a domestic setting in which other agents are family members and pets.

Young et al. describe the concept of *holistic interaction experience* as a way to analyse and design interaction between humans and robots and introduce the perspectives *visceral factors*, *social mechanics*, and *social structures* [306]. Forlizzi introduced the Product Ecology Framework, which proposes to study the social and physical context in which products such as robots are used [94]. Sung et al.'s Domestic Robot Ecology (DRE) framework is used to describe relationships with the environment engendered by the robot. The main factors are space, domestic tasks, and social actors [264].

HRI as collaboration

Moving beyond control, some authors also characterize human-robot interactions in terms of collaboration (cf. mixed initiative interaction and dynamic autonomy in [109]). The focus is on human-robot systems that are geared towards achieving joint goals. The interaction advances based on task progress and the actions of the agents involved.

The collaborative control model for robot teleoperation as proposed by Fong et al. allows for human intervention in the robot's cognitive and perceptual processes. The robot can ask for the human's assistance, enabling humans and robots to collaborate as partners [92] (see also [250, p.761]. Johnson et al. describe the Coactive System Model, which supports designing for interdependence in human-robot collaboration [149]. Hellström and Bensch propose a model of interaction that describes how humans and robots support each other's understanding through communicative actions. This can be seen as a requirement for enabling collaboration. Communicative actions by the robot seek to

[4] At an even more basic level than considering robots as socially evocative, we can also consider robots as artefacts that are constructed within a human culture. As technological artefacts, they are the result of social processes; they can be described as social facts. In addition, they can also support communication between people and in this sense function as social media [101]. This means that robots that follow a control paradigm can also encode human biases and stereotypes.

decrease the mismatch between what the human thinks the robot's state is and the robot's actual state. The model by Hellström and Bensch encompasses a double Shannon loop, with the addition of an extra transmission chain to loop information back to the sender. They add a factor for the general interaction context and note that the robot's inferences regarding the human's state-of-mind are influenced both by the human's communicative actions and the robot's state, in order to address the criticism that Shannon's model disregards context [122]. Malik and Bilberg propose a model for Human-Robot Collaboration (HRC) in the manufacturing domain, comprising the dimensions team composition, interaction levels, and safety implications [182]. De Visser et al. present the Human-Robot Team (HRT) Trust Model for the development of trust in teams comprising both humans and robots. The model describes how two actors engage in a process of (social) trust calibration. They propose that 'relationship equity' is an important aspect of trust building. This factor results from the multiple positive and negative experiences that an actor encounters over the course of the interaction history. During interaction, the actors adapt their trust stance (or attitude) towards the other agent. This trust stance helps the actors decide whether it is a good idea to collaborate on a certain task, and how to do so: is implicit agreement enough, or are formal work arrangements necessary? The trust stance is determined by the perceptions an actor has of another actor: these perceptions inform a risk assessment procedure (passive trust calibration). The process of active trust calibration involves the formation of a theory of mind on the part of each actor. This enables each actor to reason about the mental model of the other actor [64].

2.5.3 Critical discussion of existing models in HRI

In this section, we identify shortcomings in current models, which we aim to address by means of a new model in Section 2.7.

The way the interaction itself is depicted differs across the models described in Sections 2.5.1 and 2.5.2. For models that describe control relationships, the interaction is often depicted with nothing more than an arrow between the human(s) and robot(s) [92][305]. However, such a description or depiction is not informative with regards to which factors are relevant to the interaction and especially how the human is influenced by the exchange. The social and collaborative models are more detailed in this regard. The most extensive interaction component can be found in the HRT Trust Model [64]. This interaction model includes perception and collaboration components, with collaboration encompassing the subcomponents of formal work agreements, costly and beneficial relationship acts, relationship equity, and informal collaboration [64].

Something that is quite common in models of human-robot interaction or communication is that the human and robot are both depicted as agents that function in similar ways. This is especially the case in models in which we framed the human-robot interaction as collaboration. We find models in which the human and the robot are depicted as equal entities in [122] and [64], for instance. In their model, De Visser et al. assume advanced human-like social capabilities that robots may possess in the future. These supposed capabilities include possessing representations and an understanding of the behaviour of itself and other team members, and collaboration. At the moment, the authors note, the relationship is asymmetric and involves compensation on the human's part, as robots are unable to perform at this level [64]. In line with our discussion in Section 2.3.3, we want to stress this asymmetry. If interaction between humans and robots is depicted as two boxes with a double-headed arrow between them, this suggests that the agents are two individual entities with equal status in terms of agency. We argue that collaborative models seem most useful for modelling asymmetric agency, as they provide more detail regarding the interaction itself as well as cognitive factors and requirements.

Most models that frame HRI in terms of control or collaboration as discussed in Sections 2.5.1 and 2.5.2 focus on the human and the robot in the interaction. While many of these models consider the human and the robot in isolation, the models intended for interaction design take contextual and social factors into greater consideration. For instance, the Domestic Robot Ecology framework describes how the robot invites relationships with its environment [264]. A risk of viewing human-robot interactions as something strictly involving one human and one robot is that external influences and power relationships are disregarded. In practice, the interaction is influenced by outside forces, such as functionalities provided by companies. This is the case for all proprietary aspects of platforms included in the robot's hardware and software (e.g. operating system, sensors, data processing). Intelligent virtual assistants such as Amazon Alexa typically store and process text and voice commands in the cloud, and interface with other applications [45]. Functionality that depends on external processing can play an important role in the interaction between humans and robots, which becomes especially apparent in case of failure. For instance, when the company Jibo Inc. went out of business, its servers were shut down, thereby severely limiting social robot Jibo's functionality [285]. In such cases, the robot does not function by itself, but interfaces with external entities. External influences should be made explicit in order to understand the ecosystem associated with the robot. This becomes especially important over longer periods of time and when personal privacy is impacted.

2.6 How interaction and communication are conceptualized in HRI models

In this section, we critically discuss the main ways in which interaction is conceptualized in the models in Section 2.5, with a particular focus on communicative aspects of interaction. We distinguish different ways of conceptualizing interaction and communication, namely in terms of sending signals (Sec. 2.6.1), as actions in which agents implicitly construct ideas regarding the beliefs of their interaction partner (Sec. 2.6.2), interaction as joint action (Sec. 2.6.3), and interaction as a dynamic system (Sec. 2.6.4). The types of interaction discussed in this section are listed in increasing order of complexity. While discussing communication as the sending of signals can be conceived of as seeing communication in terms of discrete events, interaction in Section 2.6.2 is viewed as a chain of individual communicative actions in which agents build a conceptual model of the (mental) state of their interaction partner. These two views also bring to mind a turn-taking view of communication. Viewing interaction and communication as joint action or as a dynamic system, on the other hand, implies a more continuous view of communication, in which interaction partners can monitor each other and their environment and coordinate their actions. A joint action view of communication builds on concepts such as (or similar to) Theory of Mind and offers conceptual tools to describe an embodied coordination process between agents. The dynamic systems perspective goes one step further and integrates concepts from the first three views into a view of communication as a multimodal coordination process that is described as a self-organizing system.

Although all levels of discussion are relevant, the third view is at present most useful to describe situations in which robots are implemented to achieve shared goals in human-robot teams, which is why we focus on this view in the model we propose in Section 2.7. It is at present not easy to combine cognitive-level reasoning processes regarding common ground, for instance, with the dynamical systems approach, as noted by Dale et al. [59, p.62]. We argue that the joint action approach provides relevant conceptual tools to describe coordination, which is why this approach is the focus of the present paper. Joint action implies that communication is a participatory process in which participants have shared goals. The joint action perspective is especially useful for HRI, as it offers concepts to think about the way collaboration can be achieved.

2.6.1 Interaction and communication as sending of signals

Communication and interaction can be discussed in terms of the exchange of signals and cues. Such a discussion falls within the semiotic tradition of communication theory. Peirce and Saussure are generally recognized for their contributions to the study of signs. In Saussure's semiotics, a sign consists of the signifier and the signified. These concepts cannot be disentangled [303]. The signifier is the 'sound-image', and signified is the concept. A sound-image is the combination of what one hears and sees in response to a spoken word. Signification is the process of making use of signs with their associated meanings [176]. In Peircean semiotics, the symbol is treated as a process (semiosis) with three components, namely the sign (representamen), object and interpretant. The *sign* is the form of the symbol. What is represented by the sign is called the *object*. The *interpretant* is an effect on the person that forms the relation between sign and object [266]. Saussure noted that while the meaning of signs relies on convention, signs can also be interpreted in different ways. Signs are used in an intentional way, and for a sign to be a sign it has to be interpreted as such. Peirce, in contrast, saw signs as a means for people to think, communicate and make their environment meaningful. This does not require the sign to be part of intentional communication [303].

Typologies and taxonomies have been proposed to classify signs and cues for HRI and conversational agents. For instance, Hegel et al. [121] propose a typology of signals and cues in HRI, and distinguish between human-like and artificial signals and cues. Signals are designed to provide information, while cues (such as motor sounds) are all those features that can potentially be informative but were not necessarily explicitly designed as such. We note that the distinction between natural and artificial can cause confusion when applied to robotics. All robot signs are (at least currently) artificial. There are degrees of human-likeness (and animal-likeness, plant-likeness). Human-likeness is a spectrum, not merely a property. Feine et al. [83] propose a taxonomy of social cues for conversational agents (CA) based on a systematic literature review. This proposed taxonomy contains the main categories verbal, visual, auditory and invisible cues. The authors define a cue as *"any design feature of a CA salient to the user that presents a source of information"* [83, p.141] and a social signal as *"the conscious or subconscious interpretation of cues in the form of attributions of mental state or attitudes towards the CA"* [83, p.141]. These definitions are based on a view of cues as all those features that can provide information to the user. Cues only become social signals if the cue leads to attribution of sociality to the CA by the user. A social cue, then, is a cue that actually provokes a social reaction on the part of the user, where a social reaction is a reaction to a conversational agent that would also be appropriate if it were aimed towards another human [83]. Such a view of cues and signals is also useful for robotics, and avoids the question of whether we should regard the robot as a social agent. This view focuses the discussion on the perception of the signal as a social signal by the human.

Communication, at the level of analysing signals and cues, is used to affect the behaviour of an interaction partner or to convey information. The signal or cue is discussed as a discrete event that carries information. For instance, Hegel et al. [121] provide a table in which they note that an artificial signal such as an LED can convey activity information, for instance, while a humanlike cue such as body size conveys dominance information. While cues and signals convey information, we would like to note that context is also relevant and, as also noted by Feine et al., signals and cues do not occur in isolation [83]. Analysis at the level of signals and cues can serve as a basis for discussion, but needs to go further.

2.6.2 Interaction and communication as communicative action

Bensch et al. define interaction events as tuples of *perceived information* and *associated actions*. Interaction events link together in chains to form interaction acts [25]. Hellström and Bensch define the term *communicative action* as *“(...) an action performed by an agent, with the intention of increasing another agent’s knowledge of the first agent’s SoM"* [122, p.115], with SoM referring to the agent’s state of mind. The advantage of Hellström and Bensch’s definition is that it can be used to “computationalize" communication; it lends itself to being written in algorithmic form, in which the agent is able to compare the contents of its own belief system to the one it estimates a second agent to have regarding the first agent. While this can be advantageous for AI and robotics applications, we note that this could also be a risk if important factors in the interaction are not (or incorrectly) captured and processed. Hellström and Bensch’s definition also applies to the model they propose [122].

The concept of Theory of Mind (ToM) is central in such a view of communication processes. Krämer et al. write that *“Theory of mind (ToM) is the ability to see other entities as intentional agents, whose behavior is influenced by states, beliefs, desires etc. and the knowledge that other humans wish, feel, know or believe something"* [163, p.54]. Krämer et al. argue that the concepts of common ground, ToM and perspective taking are similar, as these concepts propose that humans have implicit knowledge regarding how other minds work, which they use as a basis for mutual comprehension, and that they enhance mutual knowledge through grounding processes [163]. In the context of robotics, De Visser et al. define Theory of Mind as *“An actor’s (e.g. actor A) estimation of another actor’s (e.g. actor B) mental model of that actor (e.g. actor A)"* [64, p.461]. There can be different levels of ToM, in which each level encapsulates the former: for example, if Level 1 contains A’s individually held beliefs, then Level 2 can contain A’s estimation of B’s beliefs, Level 3 contains A’s estimation of B’s estimation of A’s beliefs, Level 4 contains A’s estimation of B’s estimation of A’s estimation of B’s beliefs, and so on. Such a conception of Theory of Mind has some similarities with the concept of ToM for interpersonal communication, but is clearly a much reduced form. Robots do not attribute mental states to others in the sense that humans do, but may be equipped with algorithms and mechanisms to estimate emotions or beliefs held by people.

2.6.3 Interaction and communication as joint action

One perspective that is discussed in the context of HRI views communication and interaction between a human and a robot as a form of joint action. This perspective treats (linguistic) communication and coordination of actions as similar processes. Joint action can be defined as *"any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment"* [291]. The term can also be found in symbolic interactionism, which is part of the sociocultural tradition of communication research. Social acts consist of the relationship between a gesture by one actor, a response by another and a result. Blumer describes joint action as an *interlinkage* of social actions [177]. The concept has been studied in cognitive psychology and has previously been applied to interaction with artificial agents [54] and robots [128]. The current section builds on the previous one, as coordination processes involve more than awareness of the mental states of others.

Clark conceptualizes communicative acts between humans as participatory acts: the communicative action by one individual who signals and another who recognizes the signal is a joint act [49]. This view expresses that the actors are both actively involved in constructing the meaning of the information exchange. Clark discusses joint activity as an activity that is performed by two or more participants. These participants have activity roles, which helps establish a division of tasks. Participants strive to achieve (joint) public goals, but may have private goals as well. Participants in the joint activity have prior mutual knowledge or common ground, which accumulates over the course of performing the joint activity [48]. Common ground refers to mutually held beliefs, see also Section 2.8.6. Joint action involves coordinating content and process, which are themselves interrelated. Clark calls joint actions a paradoxical concept, as a group of humans does not itself intend to perform actions. Instead, individuals perform participatory actions. These actions are performed when coordination is required in order to meet common goals [47]. Clark argues that joint activity cannot be separated from language use and views language use itself as a form of joint activity. People use language to coordinate their actions, and the language used does not make sense apart from the context of action in which it is applied [48].

Krämer et al. [163] write that there are several basic communication capabilities that humans are used to and that will also be required for successful human-agent and human-robot communication, notably social perspective taking, common ground, and Theory of Mind. Coordination devices to support joint action include explicit agreement, precedent, convention, and joint salience. Whether something is salient to all participants in an interaction, depends on their common ground. Participants can usually assume the solvability and sufficiency (with respect to the available information) of a current coordination problem, when one of the participants proposes the problem themselves [47]. Vesper et al. [291] discuss coordination mechanisms between humans in intentional joint action. Participants in the joint action use mental representations of the joint action goal and the task in order to monitor task progress. The co-actors share information throughout the task using mechanisms such as shared gaze, predicting the actions of the other actors involved, sensorimotor communication, and haptic coupling. They also express emotions and interpret those of others. The mechanisms they use for joint actions are coordination smoothers (e.g. synchronizing actions), communicated and perceived affordances, and cultural norms and conventions [291]. Mutlu et al. [193] describe coordination mechanisms established in cognitive science, such as joint attention, action observation, task sharing, coordination of actions, and perception of agency. Mechanisms such as gaze, action observation and conversational repair have also been investigated in the context of robotics [193].

Clodic et al. [51] present a framework for interaction between humans and autonomous robots with the aim of achieving human-robot joint action. The authors align Pacherie's theory of joint action with a three-layered robot architecture. Clodic et al. write that joint action not only requires individual agents to have common goals and be able to execute plans and actions, but that they must also be able to coordinate their individual plans. This coordination of subplans needs to occur prior to as well as during execution. This requires the capacity to monitor and predict the actions (and intentions) of one's partner. They note that motivational uncertainty (do the goals of the other agent align with mine?), instrumental uncertainty (how do we achieve the goal?) and common ground uncertainty (are we both on the same page regarding the goal and actions to take to get there?) can negatively affect mutual predictability. The authors note that self-other distinction would be a required capability for the robot: it would have to maintain 'mental' models of both itself and of the human [51]. Compare this to the concept of Theory of Mind as described in Section 2.6.2.

2.6.4 Interaction and communication as a dynamic system

Interaction can also be conceived of in a different way, namely as a dynamic system in which the interaction partners are simultaneously monitoring each other and their environment for cues and signals. Dale et al. [59] propose dynamical systems theory as a framework for a more comprehensive theory of human interaction. The authors write that many theories and concepts have been proposed to describe aspects of human interaction, such as perspective taking, joint action, ToM, and mimicry. In order to understand how these different accounts and types of processes form a multimodal coordination process, they propose that human interaction functions as a synergetic, self-organizing system. Sandry [240] refers to human-animal communication to argue that communication and interaction are more like a dynamic system than a dialogue with strict turn-taking. She writes: *"Communication operates as a dynamic system during this type of embodied communicative situation, and signals between communicators overlap as human and [animal] continually reassess each other's position, perceived intention and likely subsequent action"* [240, p.40]. In this dynamic system, sometimes the meaning of an individual communicative act can be understood easily, while in other cases the meaning can only be derived from other communicative acts in context. This can also be applied to robotics: consider the situation of monitoring a robotic system in a manufacturing context. The interaction does not follow a script, but rather consists of the human paying attention to the system, with events (such as being alerted by the robot if something goes wrong) provoking action on the human side.

The idea of interaction as a dynamic system is echoed in the concept of interaction fluency. Hoffman [127] proposed fluency metrics to assess how fluently a human-robot team interacts. By making an interaction more fluent, it is proposed, one moves away from a strict turn-taking interaction towards an interaction in which the robot starts to anticipate on human actions, allowing for overlap between human and robot actions instead of only one agent acting at a time. Conceiving of interaction as such a dynamic system is one step beyond mixed-initiative interaction, as it not only involves considering who takes initiative: it also requires anticipating the other agent's actions.

2.7 An Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI)

In this section, we propose an Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI). Humans and robots are very different entities. To model interaction or communication between these different types of entities, they should be depicted in an asymmetric way. Robots can also be subject to outside control (external influence). A model of human-robot interactions in which the differences in functioning between human and robot agents are highlighted, can help identify possible mismatches between robot capabilities and human expectations. Based on the discussion of the models of communication and interaction between humans and robots found in the literature, we propose that a model for collaboration between humans and robots that depicts the human and the robot as different types of agents (thus, an asymmetric model) will be useful for identifying how the human and the robot operate differently, and predicting mismatches between robot capabilities and human expectations of those capabilities. A joint action perspective on interaction emphasizes the collaborative co-construction of meaning by the agents involved, and is more suited to a scenario of human-robot collaboration in which the robot supports human work, which is why we chose to use this perspective on interaction from Section 2.6.

The model as depicted in **Figure 2.1** shows a robot architecture with common components and depicts the processes by which a human and a robot may establish common ground. This model will be used as a basis to compare the human and the robot sides of the model in Section 2.8. Moreover, based on the identified differences, we propose design recommendations in connection with a version of the model, as illustrated in Figure 2.3. The models and design recommendations are mainly intended for interaction designers in the field of HRI.

2.7.1 The model

We propose a model based on Kincaid's diagram of the components of the convergence model of communication [154], Christensen and Hager's diagram of the robot sensing process [250, Ch.4] and Bauer et al.'s diagram of mechanisms for robot joint action [24], see **Figure 2.1**. We draw a parallel between theory regarding joint action and Kincaid's convergence model of communication. Kincaid views communication as a process that has the aim of arriving at mutual understanding, which is achieved through creating and sharing information. His model contains components such as mutual agreement and collective action. The terms used by Kincaid are similar to the vocabulary in the literature on common ground [5]. Here, we use the term *joint action* in line with the HRI literature. Kin-

[5] The combination of the terms 'mutual understanding' and 'mutual agreement' in Kincaid are similar to the term common ground, while Kincaid's term 'collective action' is similar to the term joint action. Kincaid defines mutual understanding as "the combination of each individual's estimate of the other's meaning which overlaps with the other's actual meaning. In other words, mutual understanding is a combination of the accuracy of each individual's estimate of the other's actual meaning" [154, p.32]. Kincaid describes mutual agreement the following way: "When two or more individuals believe that the same statements are valid, they become true by consensus, or mutual agreement with some degree of mutual understanding" [154, p.31] while Clark describes common ground as "...The sum of (...) mutual, common, or joint knowledge, beliefs, and suppositions" [46, p.93]. Both the definitions by Kincaid and the one by Clark refer to mutually held beliefs. We refer to this concept as 'common ground'. Collective action (Kincaid) is the "(...) result of the activities

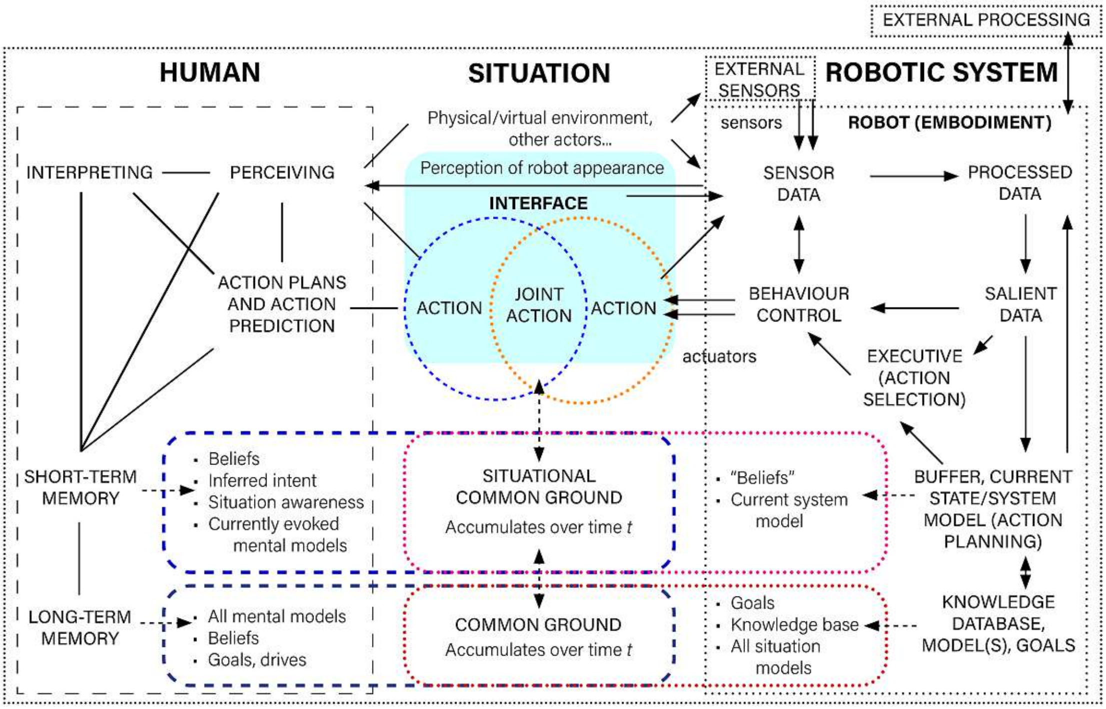


Figure 2.1: Here we propose an Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI). This is a model with global similarity between the components/organization on the human side and the robot side, which allows for comparison. However, the processes that occur are different, to allow for identification of differences between the two agents. The 'interface' includes everything that allows for communication/information exchange, including observed actions and observations of embodiment. **On the human side:** Interpretation of percepts leads to new beliefs, which are added to short-term memory. When actions are executed, their effect is predicted and the action execution is monitored. **On the robot side:** Moving from data to processed data requires feature extraction and data processing. Processed data is then further processed; algorithms determine which information is salient through attention mechanisms, data fusion, and matching processed data to patterns in the database. Between the buffer and the knowledge base, data updating and retrieval processes occur. Solid-line arrows indicate processes, dotted arrows indicate a theoretical connection.

caid's concept of mutual understanding can be related to the model and definitions by Hellström and Bensch, as well as the concept of ToM as used in the robotics and AI literature (see Sections 2.6.2 and 2.6.3).

The proposed model in **Figure 2.1** includes indications of processes that could occur in communication in human-robot teams. However, we note that the model of the human side is highly abstracted and incomplete. We do not propose that this is an accurate, complete depiction of human processing,

of two or more individuals (A and B), built upon a foundation of mutual agreement and understanding" [154, p.31], while joint action (Clark) is activity performed by two or more individuals based on common ground and joint action goals.

but we still include a sketch of the processes on the human side to allow for comparison between the human and the robotic agent. The model in **Figure 2.1** is structured in such a way that both agents are as similar as possible on a high level, while still indicating differences in terms of processes. The model visually expresses the asymmetry between humans and robots. The robot side displays processes that may be lacking in many robots. The point is that the model depicts the human and the robot with respect to similar processes, allowing for comparison, while still being expressive of major differences between the agents that are relevant for interaction design in HRI. (If the model were completely asymmetric, with no correspondence of components or processes on the human side with components on the robot side, the two sides could not be compared. The asymmetry arises partly by means of comparing the two sides.)

We now discuss the components listed in the 'human' component of the model, followed by a discussion of the situation component and the processes occurring on the robot side.

2.7.2 Discussion of the proposed model: Human

The processes on the human side of the model are mainly intended for illustration, drawing parallels, and illustrating contrasts with processes occurring on the robot side. The processes on the human side are based on the model of human-human communication by Kincaid and Endsley's models of situation awareness, which models human-technology interaction. Endsley proposes a model of situation awareness (SA) and discusses its role in the context of human decision-making. SA plays a role in applications ranging from air traffic control and tactical systems to decision-making in everyday activities. Endsley distinguishes three levels of situation awareness: (1) perceiving information, as well as (2) comprehension/understanding, and (3) forecasting future states to aid decision-making. Endsley proposes a model regarding the interrelation between a person's goals, mental models, and situation awareness, as well as a model depicting the mechanisms that are important in establishing situation awareness [79]. The components long-term memory, short-term memory, and the terms situation awareness and mental models are derived from her models. The long-term memory component expresses that the human does not have all their knowledge and experience at hand to apply to a new situation, but instead retrieves a subset of knowledge in response to the current situation.

As mentioned earlier, Pickering and Garrod [214] proposed that human comprehension and production processes should be understood as tightly interwoven, which supports human agents in predicting the actions of their interaction partner as well as their own actions, interweaving actions within a coordination process, and achieving joint action. Separating production and comprehension would lead to a model that Pickering and Garrod [214] refer to as a "cognitive sandwich" (as coined by Hurley [136]) in which cognition is sandwiched between perception and action. While this "cognitive sandwich" may well exist in many robot architectures, we assume in the model that perception, action and interpretation are tightly coupled on the human side, as denoted by lines (rather than the arrows indicating one-way processing on the robot side). In contrast, the processes on the human side are not one-way but can mutually influence each other.

When human behaviour or human cognitive processes are modelled, these models are necessarily limited; the model developer is required to make assumptions regarding human behaviour, perception and cognition. While models can be useful, as in the context of this paper, it remains important not to lose sight of people's embodied, individual and varied experience. It should be noted that every human is different, has individual needs and abilities as well as a personal background and identity. This variety and these differences need to be kept in mind rather than assuming an imaginary 'general human'. Spiel argues that we should *"appreciat[e] the plurality of human bodies instead of assuming a specific embodiment"* [256, p.2] in the context of technology design for embodied interaction. The same goes for human cognitive processing, action and perception capabilities, and other components of the model in Figure 2.1. The components listed on the human side are a limited selection and real-world human capabilities go beyond what is included in the model, as they are far more complex and more varied than what can be captured in an illustration.

2.7.3 Discussion of the proposed model: Robot

The model on the robot side is based on a model of the robot sensing process [250, Ch.4], a model of mechanisms enabling joint action for robots [24] and theory on cognitive architectures. For a discussion of how cognitive architectures can be structured, we refer to the survey by Kotseruba and Tsotsos [159] and the review by Chong et al., who compare the cognitive architectures SOAR, ACT-R, ICARUS, Beliefs-Desire-Intention (BDI), the subsumption architecture and CLARION [44]. The goal of research on cognitive architectures is to model the human mind and achieve human-like intelligence. Cognitive architectures are also developed and implemented for robotic systems, see for instance [274].

With regards to the model proposed in Figure 2.1, cognitive architectures and computational models of human(like) behaviour are useful both for developing the system architecture on the robot side of the model as well as for maintaining representations of the state of human interaction partners in the robot's memory, in order to facilitate collaboration. Perception, attention, action selection, memory, learning, reasoning, and meta-cognition are the main features of cognitive systems (and have been developed further than, for instance, emotion and creativity) [159], and these are the features we focus on in the proposed model. Based on the research on cognitive architectures, we included short-term and long-term memory components (see, for instance, the SOAR architecture in [44]). In the next sections, we provide more details regarding robot sensing, knowledge representations, reasoning, learning, action selection and actuation. The discussion will focus on the robot's awareness of humans and their behaviour, as this is relevant in a communication process.

Sensing

In the model in Figure 2.1, the sensing process consists of the components *Sensors* and *External Sensors* that collect sensor data, the component *Processed Data,* and the component *Salient Data*. The sensing and perception process in robotics has been described as a process that involves updating an existing partial world model with sensor data. This process involves feature extraction, matching or associating data with an already existing model, updating and integrating new knowledge in the model, and prediction of future states (which influences data matching) [250, p.88]. Yan et al. [304] identify feature extraction, dimensionality reduction and semantic understanding as key components to social robot perception systems.

According to Christensen and Hager [250, Ch.4], sensors for robotics applications can be classified in the following way: tactile, haptic, motor/axis, heading, beacon based, ranging, speed/motion, and identification. The types of sensors installed on a robotic system depend on the application; for instance, medical robots and industrial robots will need different sensors than assistive robots intended for social interaction. We can make a distinction between sensors for interoception and exteroception. Interoception refers to sensing the robot's state (e.g. motor currents). Exteroception refers to sensing the external world (e.g. distance to an object) [250, Ch.4]. Sensors can also be classified as passive (does not emit energy) or active (emits energy in order to sense) [250, p.452]. They can also be on-board or external (consider, for instance, the concept of the Internet of Things; in such a scenario, the robot may have access to external sensors and devices).

The main types of signals that social robots make use of for interaction with humans are based on visual, audio, and tactile interaction modalities, as well as ranging sensors [304]. Visual-based signals can be captured using 2D and 3D cameras (using depth information with RGB-D cameras or stereo vision). Audio data can be captured using microphones for subsequent speech recognition, another key aspect in human-robot interactions [275]. Data from different sensors and interaction modalities (vision, audio, touch) are combined and subjected to processing (for instance, by means of computer vision methods using Hidden Markov Models (HMMs)) [275]. In order to make sense of high-dimensional data, statistical techniques (such as principal component analysis) can be used for processing the data and extracting features from the data in a lower-dimensional feature space, after which the data can be used for applications such as object recognition [304]. Recognizing humans and features of humans and their behaviours are of high importance for social interaction with robots. Research on computational HRI has focused on topics such as detection of body pose, face recognition, activity and gesture recognition, and interaction engagement [275].

The *Salient Data* component is included in the model to indicate that processed data may be further processed to identify the features in the environment that are most relevant to the interaction. Unimodal feature extraction is often not robust enough; therefore, multimodal feature extraction methods can be used, in which data from separate modalities are combined into a saliency map [304] or measures of object saliency (e.g. [274]).

Knowledge representations, reasoning and learning

A suitable knowledge representation is required for reasoning about information and storing it. The field of knowledge representation is concerned with finding representations that are adequate in an epistemological sense (represent referents in the environment in a compact, precise way) and in a computational sense (that is, efficient) [250, Ch.9]. The formalisms used for knowledge representations and making inferences are mainly based on logic and probability theory [250, Ch.9]. Reasoning has specific issues in robotics applications as compared to other types of knowledge-based systems. Robots are embedded in dynamic environments and have to interpret and respond to environmental information (partially) autonomously in near real-time. Approaches that try to remedy these issues include fuzzy logic approaches and embedding time constraints within the robot's architectural design [250, Ch.9] (see also the KnowRob system for an example [269]). Learning on the robot side can occur in different ways. Kotseruba and Tsotsos describe learning as *"the capability of a system to improve its performance over time"* [159, p.50], based on experience. They distinguish between declarative and non-declarative learning, where non-declarative learning encompasses the learning mechanisms perceptual, procedural, associative and non-associative learning [159]. One specific type of AI is machine learning. Hertzberg and Chatila define machine learning in the context of robotics as *"the ability to improve the system's own performance or knowledge based on its experience"* [250, p.219]. Methods include inductive logic programming, statistical learning, and reinforcement learning. Learning can be supervised or unsupervised [250, Ch.9].

A robot architecture can also be designed to support some level of metacognition. Metacognition includes introspective monitoring of the robot's status and processing (e.g. self-observation) and Theory of Mind (ToM) [159]. In order to accommodate social interaction with humans, robots can be equipped with mechanisms based on ToM (which means, in the context of cognitive architectures, that the system infers others' mental states and uses this information for decision-making [159]) and ways to explicitly model humans and human behaviour. Cognitive architectures have been proposed that draw on the concept of ToM, in order to infer human intentions from goal-directed action [275]. However, most social robots are far from full ToM. At present, research has been conducted on the development of capabilities such as parsing human attention, which may aid in the achievement of human-robot joint attention, and predicting human action in order to be able to anticipate on it [275]. Hiatt et al. [123] review different ways of modelling human behaviour that can be implemented in a robotic system with the aim of enabling the robot to understand a human teammate's behaviour. They write that computational approaches (such as conventional machine learning approaches) can be useful in situations in which rational, 'ideal' or 'typical' performance by humans can be assumed, but this leaves little room for human error or deviation from set norms, although such deviations are to be expected in human-robot collaboration. They also discuss computational/algorithmic approaches such as HMMs and the cognitive architecture ACT-R/E [123].

Action selection and actuation

Action selection can occur dynamically (choosing one option from a set of alternatives) or in the form of action planning (as is common in traditional AI) [159]. Planning problems are usually described as sets of states with actions that can induce transitions between states. The goal is to find a suitable series of actions from the start to the goal state. Action planning can involve working towards a common goal for efficient human-robot collaboration [24]. Robot planning uses planning methods that make use of formalisms from logic and probability theory to complement motion planning [250, p.219]. In the research area of computational HRI, fluent meshing of actions, human-aware motion planning, object handovers, and collaborative manipulation are important research foci for robot action planning [275].

Motion trajectories by the robot should be possible to execute; therefore, motion planning needs to take the robot's kinematic constraints into account. Aside from achieving task goals, robot actions such as robot motion can communicate intent to an interaction partner or observer, whether or not the action is planned to be communicative. Motion can also have a communicative aspect: instead of purely functional motion planning, generating motion that is legible and/or predictable to human interaction partners can also be considered [70]. Social robot navigation has a social component as well, as demonstrated by the research topics of approaching humans, navigating alongside people, and human-aware robot navigation [162][229][275]. The use of gestures and gaze cues, proxemics, haptics, affect, emotions, and facial expressions have been studied as nonverbal behaviour that can be implemented in robots for communication [275]. The robot can also use other interaction modalities as part of a communication process, for instance by making use of auditory signals (see also Section 2.6.1) or changing the state of a graphical user interface that is part of the system.

2.7.4 Discussion of the proposed model: Situation and Interaction

In the model, the *Situation* refers to the current proximate physical and social environment (the interaction context), the current constellation of agents, objects and environment, close to each other in space and time. It includes other agents or actors that may be involved or referenced in the communication process.

Joint action consists of actions involving both agents that have the aim of establishing common ground or achieving shared goals. Joint action is a subset of all actions, including those actions that advance the human and the robot in their joint action goal. *Situational common ground* is the subset of the interaction partners' beliefs and goals that are shared in the current situation. We included the component situational common ground as something separate from common ground, based on Endsley's theory on situation awareness, which holds that not all information is in consciousness; this is true only for a subset of information and mental models.

Joint actions are a subset of all actions carried out by the participants in the interaction. These actions are the components of larger joint activities (cf. Clark, [48]) and move the participants closer to a desired goal state. Clark differentiates between a joint act and a joint action. The former is discontinuous, while the latter is a continuous coordination process. Clark distinguishes phases as the distinctive elements that make up joint actions and that allow them to be coordinated, defining phases as *"a stretch of joint action with a unified function and identifiable entry and exit times"* [47, p.83]. Examples of joint actions are giving a person a handshake or asking someone a question. In the proposed model, no distinction is made between actions and communicative actions. However, a detailed look at research in the semiotic tradition and the work that has been done in HRI on classifying signs and cues can be useful to specify the communicative aspects of actions further.

Clodic et al. identify three levels of uncertainty, namely instrumental uncertainty (related to joint action), common ground uncertainty (related to common ground) and motivational uncertainty [51]. We can identify these levels of uncertainty in the model. Instrumental uncertainty occurs on the levels of action and situated common ground. Common ground uncertainty and motivational uncertainty both occur on the levels of situational common ground and common ground. The robot does not have 'personal' goals. This may result in increased motivational uncertainty on the human side regarding the intentions of the developer of the robot, its software, or owner, if the motivations/goals of the robot developer are not communicated.

Humans and robots can only have reduced common ground as compared to the common ground shared by humans. If the robot can only sense and act, the common ground factor in the model would become irrelevant, and instead of "joint action", we might label the aggregate of human and robot action as a "collection of actions" instead.

Participants in the interaction have internal goals or goals that have been defined externally. Participants are trying to achieve goals while engaging in joint activity, most notably the *domain goal* in Clark's words, yet participants can also have procedural goals, interpersonal goals and private agendas [48, p.34]. In human-human communication, high-level goals are usually internally defined (and then possibly negotiated), but this is not the case for robots. High-level goals may be externally dictated by human interaction partners or the company or companies that produced the robot and its components. Subgoals, on the other hand, might be either external or internal, derived from high-level goals (e.g. moving to intermediate location B while moving from A to C).

Having a 'joint intention' or a common goal refers to a joint, participatory aim that is shared across participants, *"a joint commitment to perform a collective action while in a certain shared mental state"* in the words of Cohen and Levesque [54]. The notion of joint goals, or working towards achieving a common goal, is not necessarily useful in all cases, especially if the robot is intended for social interaction and/or operating in a (semi-)public space. For instance, if a (human) visitor to a conference approaches a humanoid robot and starts waving in front of it and muttering phrases to it to see if it will respond, this behaviour could be said to have a goal on the human side (even if subconscious), namely to entertain themselves and figure out what the robot can do, but it cannot really be said to constitute collaboration or 'working together'. Joint action arises only when the action is acknowledged or responded to by another agent, and the goals of both agents align. The robot's high-level goals are defined externally, but lower-level goals (such as moving to intermediate location B while moving from A to C) can be defined internally.

One may read the agents as acting in a very goal-directed way on the basis of the preceding text (in the way of Saussure, instead of Peirce). However, a view of actions and signals as supporting a process of reflection is not excluded, and actions and communication can also be viewed in the model as a means of thinking. For instance, consider a case in which a robot pushes over a stack of blocks repeatedly and observes what happens.

2.7.5 Practical example

In this section, we walk through the model using the practical example of lexicon learning. We will shortly elaborate on the example. The human teaches the robot new words by pointing to objects on a table and naming the objects. The robot stores representations of the object and the words the human uses to name those representations (accumulation of common ground). After the teaching phase, the human asks the robot to name the objects on the table that the human points at (joint action). In this fictive example, consider the robot to have a moveable head and to have pointing detection, speech recognition, basic object recognition, and face recognition functionality.

We work this out in a script form in which each robot action is specified. Only human perceptions, thoughts and actions are included. We do not presume to guess the human's inner workings, but propose one possible option for what the human may infer based on robot actions and other events. This depends also on other factors, for instance whether the human is an expert user or a novice. In practice, the expectations of (multiple different) human interaction partners can be elicited in the context of interaction experiments by means of methods such as think-aloud and post-experimental questionnaires or interviews. With regards to the interaction component, common ground uncertainty is included at relevant points. Note that we discuss one action-response pair, so a single joint action.

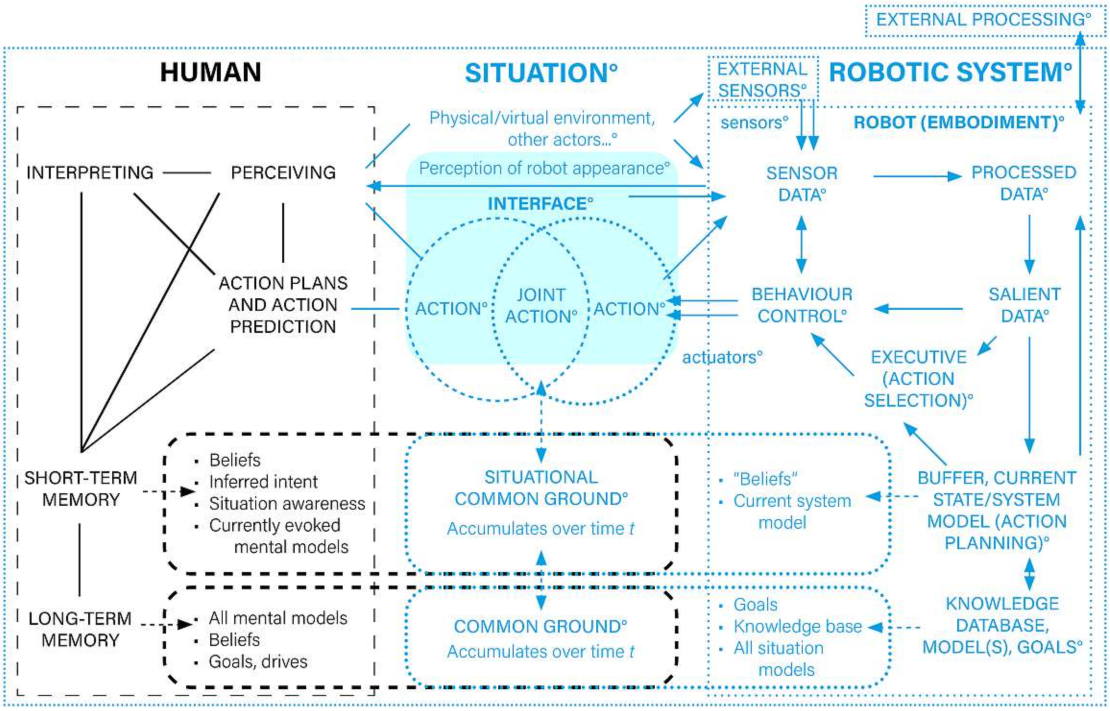


Figure 2.2: Marked in blue: Possibilities to influence. It is possible to directly influence all components and processes on the robot side by reprogramming it. However, it is not possible to directly influence the human side, except on the level of input/output: it is possible to give the human different information or influence the way the human is able to execute actions. Note that it is possible to influence human processing indirectly, as things such as hormones, food, and attention all influence the way processes on the human side operate. It is also possible to change a person's belief system by changing the information environment they are exposed to.

What can be useful about working out such a script, is that it forces the developer to be very specific regarding expected human thoughts and actions, which yields hypotheses that can be tested. It can also help with identifying whether the robot's behaviour needs to be modified.

The items labelled **Human (X)** indicate an action or process on the human side, while the items labelled **Robot (X)** indicate an action or process on the robot side. They are presented here in a sequential way, although some actions that are listed as sequential can also co-occur at the same moment. The italicized items marked with quotes are thoughts or verbalized human thoughts, depending on whether the items were elicited by means of brainstorming by the researchers (as they are in this case) or the method of think-aloud.

Human common ground uncertainty (1) Uncertainty on the human side before naming the cup

- Instrumental uncertainty: *"What can the robot do? How will the robot act?"*
- Common ground uncertainty: *"What does the robot know?"*

- Motivational uncertainty: *“Does the robot have the goal of learning the names of these objects? ... I suppose so, the researcher told me?"*

Human (1) Perception

- *“I see a yellow robot with large eyes, a torso and two arms"*

Human (2) Interpretation

- *“Cute! I guess it is looking at me. I wonder what it can do."*

Human (3) Action

- The human points at a cup. The human pronounces the word *“cup"*.

Robot (1) Data is captured by the robot’s sensors

- Audio is captured by the microphone.
- Image data is captured by the robot’s camera.
- Depth information is captured by the robot’s depth sensors.

Robot (2) Feature extraction

- A pre-trained image processing algorithm identifies that there are three objects in the camera view that do not have a stored label associated with them. A pre-trained object recognition algorithm recognizes a human face and a pointing hand.
- Speech recognition software recognizes the word *“cup"*.
- The direction in which the hand is pointing is inferred using the video and depth information, and stored as the approximate pixel area on the video image.

Robot (3) Attention and data fusion

- The object **[object1]** that the human pointed at is inferred.
- The location of the human face is inferred.
- The speech recognition result *‘cup’* (semantic label) is associated with the pixels from the image that were labelled **[object1]**.

Robot (4) Action selection

- The robot looks in the direction of the object that the human is pointing at.

Human (4) Perception

- *“The robot is looking at the object"*

Robot (5) Buffer: storing data in short-term memory

- The semantic label *‘cup’* and **[object1]** are placed in the buffer.

- The location of the human face is stored.

Robot (6) Matching: storing data in long-term memory

- **[object1]** and *'cup'* are stored in long-term memory.

Robot (7) Action selection based on successful storage of item in short-term memory

- After a delay of 2 seconds, the robot moves its head in the direction of the human face.

Human common ground uncertainty (2) Uncertainty on the human side after naming the cup

- Instrumental uncertainty: the robot acknowledged the human's action when it looked at the object. *"The robot looked at the cup when I pointed at it, so it must have noticed what I pointed at."*
- Common ground uncertainty: *"Did the robot understand that the object is called a cup? Does the robot already know that the object is a cup?"*
- Motivational uncertainty: *"Is the robot currently trying to infer the object name?"*

After this interaction, the common ground between human and robot can be constructed as follows:

Beliefs Human (ToM Level 1) Human knows *"cup"* is associated with the object cup.

Beliefs Robot (ToM Level 1) Robot inferred that *'cup'* is associated with **[object1]**.

Beliefs Human (ToM Level 2) The human does not know if the robot knows the object is a cup.

Beliefs Robot (ToM Level 2) Robot inferred that human calls **[object1]** *'cup'*.

Beliefs Human (ToM Level 3) The human does not know if the robot knows that the human does not know if the robot understood it is a cup.

Beliefs Robot (ToM Level 3) The robot did not acknowledge that the object is a *'cup'*, so the robot may infer that the human does not know that the robot knows **[object1]** is a *'cup'*.

Situational common ground (after interaction) Both human and robot associate the object with similar labels (*"cup"*/*'cup'*), but the knowledge that the robot has associated the object with the label is not common ground. The robot should use this information to communicate that the word *'cup'* is now common ground, or confirm otherwise.

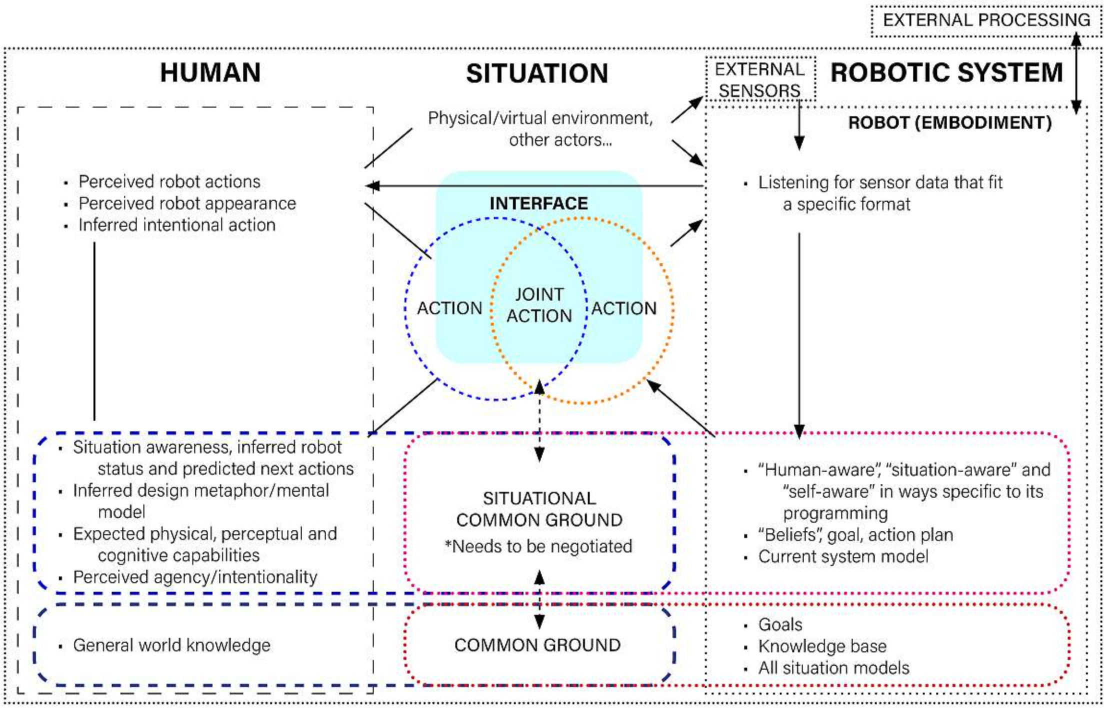

EXTERNAL PROCESSING
HUMAN
SITUATION
EXTERNAL SENSORS
ROBOTIC SYSTEM
Physical/virtual environment, other actors...
ROBOT (EMBODIMENT)
• Perceived robot actions
• Perceived robot appearance
• Inferred intentional action
INTERFACE
• Listening for sensor data that fit a specific format
ACTION
JOINT ACTION
ACTION
• Situation awareness, inferred robot status and predicted next actions
• Inferred design metaphor/mental model
• Expected physical, perceptual and cognitive capabilities
• Perceived agency/intentionality
SITUATIONAL COMMON GROUND
*Needs to be negotiated
• "Human-aware", "situation-aware" and "self-aware" in ways specific to its programming
• "Beliefs", goal, action plan
• Current system model
• General world knowledge
COMMON GROUND
• Goals
• Knowledge base
• All situation models

Figure 2.3: This figure is based on the architecture in Figure 2.1. It is a simplified version of the model that is meant to clarify the model's relation to the design recommendations.

2.8 Design recommendations

Norman identifies seven design principles for interaction design, namely discoverability, feedback, a conceptual model, affordances, signifiers, mappings and constraints [199]. These principles point to the importance of making sure a person interacting with a product or interface is able to determine which actions are currently possible, what the state of the device is, making sure that the person has a good conceptual model of how the device operates, and giving feedback in response to the person's actions. Such design principles have also been proposed for HRI [72][282]. Here, we propose several design recommendations based on the AMODAL-HRI model for interaction designers in the field of HRI.

The human and the robot are different types of entities. In the model in Figure 2.1, they are depicted in an abstracted form, with similar-yet-different processes and components. This allows for comparing the two sides. Such a comparison can help identify potential communication failures. Differences in the capabilities of the robot versus the human can lead to communication failures if the human has expectations regarding robot behaviour that the robot does not meet. In Section 2.8.1, we discuss some of these differences, such as differences in perceptual capabilities.

In Sections 2.8.2-2.8.6, we discuss design recommendations that are aimed at overcoming or ameliorating these differences. In order to better illustrate the design recommendations and the factors that are most relevant in the communication process, we have added an additional version of the model in Figure 2.3. The way the robot processes information is still the same as in the robot architecture version in Figure 2.1.

The design recommendations in this section are related to the concept of transparency (and related themes such as explainability, understandability, and interpretability). It has been proposed that transparency can aid a user when it comes to understanding how an AI system works and performs decision-making processes. Transparency for robotics and AI means that the system informs the user regarding what the system is doing and why, (1) making it easier for the end user to predict the system's future actions and (2) in order to enhance the user's trust in the system. The function of transparency is to support the end user in understanding the reasons behind a system's decisions and actions, and it helps the user check if the system is working correctly [84].

2.8.1 Differences between the human and the robot

Before diving into a discussion of high-level differences between humans and robots, it should be noted that there is an incredible diversity when it comes to human bodies and physical abilities, how humans interact with technologies, a diversity of cognitive abilities, and so forth. Robots should be designed in ways that ensure accessibility in order to accommodate a diverse group of end users. While there is a large variety in robot morphologies and the ways robots can be programmed, robots are far more generic, and the way one robotic system functions can in theory be replicated on (similar) robotic systems.

Robots and humans have different perceptual capabilities, as there is a difference between robot sensors and human sensing. Robot sensors have a different range and may perceive different information, compared to the human sensory organs. For example, the camera view may only cover a fraction of the range and area that a human eye can perceive. Liu and Chai write that humans and robots lack a shared perceptual basis due to differences in perception and reasoning. The robot's perceptual capabilities are limited and markedly different from those of the human interaction partner due to, among other things, its specific computer vision algorithms (one particular problem that is well-researched is the referential grounding problem, which in the context of HRI refers to the problem of connecting references by a human interaction partner to objects in the environment to perceptions of those objects by a robot) [178]. In addition to differences introduced by the particularities of a machine learning algorithm, there are differences in sensor reliability in different conditions [250, p.103], and different modelling conventions can be chosen to represent the world (environment) and the state of the robot based on sensor data [250, p.104]. On the other hand, a robot can have additional sensors, such as infrared sensors, and obtain information a human does not (biologically) have access to. This can lead to the problem that it can be difficult for a human to understand what the robot is (in)capable of doing or perceiving. The robot's capabilities and functionality can be conveyed to the human through training [51] (different instruction methods such as video tutorials are possible, as is trial-and-error exploration [37]). Others propose that the robot could assess the reliability of its perceptions through human-robot dialogue [178].

Another difficulty with respect to robot perceptual and cognitive capabilities concerns asymmetries in the recognition versus production of speech. Thomaz et al. note that it is much more challenging to make robots capable of recognizing speech than it is to make them capable of producing speech with a similar level of complexity. However, if robots are capable of producing speech at a certain level of complexity, this may lead people to infer that the robot will be able to understand their speech at that level of complexity [275].

In addition, robot perception and reasoning can be biased due to biases in the datasets that were used to train machine learning models. For instance, with regards to gender bias, Wang et al. write that in dataset COCO, *“images of plates contain significantly more women than men. If a model predicts that a plate is in the image, we can infer there is likely a woman as well. We refer to this notion as leakage"* [295, p.5310]. If robots make use of machine learning algorithms that were trained using such datasets, they may amplify and reinforce existing societal biases. While humans are certainly biased as well, our biases are not amplified in a similar way.

With regards to physical movement, humans and robots have different action capabilities: they have a different workspace, and different possibilities regarding motion speed and acceleration. (Again, we note that there are differences among humans as well.) Humans and robots have different action possibilities and communication modalities (for instance, a robot may be able to communicate using LED lights) due to differences in embodiment and morphology. Hoffman et al. write that there is a large variety in robot morphologies. For instance, the robot’s embodiment can be zoomorphic or humanoid, some are able to manipulate objects with arms while other have wheels, et cetera. They write that perceived robot morphology influences the capabilities that humans expect a robot to have [132].

It can be expected that humans will adjust more easily to robot limitations than vice versa. Dale et al. [59] describe that in studies of computer-mediated communication, people accommodate to their interaction partner when they think their interaction partner is simulated, for instance by using less complex language and by taking the other’s perspective more often, as they aim for a maximum level of mutual understanding. While it may not be desirable to rely on humans to adjust to the limitations of machines as a design strategy, we mention it here as a difference.

2.8.2 Affordances

Signifying affordances is useful in the context of communicating differences in robot sensing and action capabilities to the human. The term affordance, coined by Gibson, refers to an agent’s possibilities for action when interacting with an object or environment [199]. This agent could be a human or a robot. The affordance concept can be applied to a human’s possibilities for action when interacting with an object or device such as a robot. In this case, signifying the affordances (through signifiers) in order to make them discoverable is a human-centered design challenge (see [88]). The concepts of affordances and signifiers are closely tied to Norman’s concept of discoverability, which is a human-centric notion that indicates that the human can find out how the device functions through directed experimentation or applying mental models from similar devices.

For example, if the robot has the capacity to record audio and interpret speech, the robot affords being spoken to by a human interaction partner. Affordances can be communicated to the human interaction partner by using signifiers. That is, by communicating such things as what the robot can read with its sensors, it can be communicated to the human how the robot can be interacted with. The robot’s actions can give the human clues regarding the actions that are possible with the robot.

For instance, if this particular robot can only understand the words “yes" or “no", it can signify this by asking the participant to answer with “yes" or “no" after each question.

Another consideration regards embodiment design, for instance with respect to the placement of sensors or how sensor placement is communicated. Humanoid robots are often designed to give the impression of having eyes, but sensors that capture image data are not necessarily placed at the same location. The same goes for microphone and loudspeaker locations. For instance, on the Pepper robot, the microphones are placed on top of the head, while the speakers are at the ear locations [253]. When it comes to signifying affordances, if a humanoid design is chosen, it may be desirable to place sensors in a way that corresponds to the approximate location of human sensory organs. If the sensor placement deviates from expectations significantly, this may need to be communicated to human interaction partners.

Regarding the example in Section 2.7.5, we can identify various opportunities for communicating the state of the teaching and coordination process by applying the concept of affordances. First of all, the robot can indicate that it can be talked to by asking questions regarding objects in the environment or through an attentive posture (affordances/signifiers). Secondly, the robot has already indicated that it can focus its attention and attend to items the human talks about by looking at the object the human is pointing at (see label **Robot (4)**).

Design Recommendation 1 Signify affordances to indicate how the robot can be interacted with.

Design Recommendation 2 Communicate the robot’s intended function and action possibilities in different situations to novice users, so the human can understand how the robot functions.

See also Fischer [88], who argues that the robot can communicate affordances implicitly by means of leading questions, and who argues that we can make use of the “downward evidence" signalling strategy: when humans are presented with high-level capabilities, they expect the robot to have lower-level capabilities as well. This is an implicit way of signalling affordances [88].

2.8.3 Mental models and design metaphors

Mental models have been defined as *“the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states"* [236, p.7]. Interpretation and understanding in humans can be described as occurring through the application of certain frames or mental models [79] to a situation, based on previously experienced situations[6]. In HRI, the mental model concept has

[6] Endsley relates the concept of the mental model to that of the situation(al) model, a way of understanding the current state of a system. A situation model/mental model can be used to identify critical cues and elements to attend to, understand the meaning of elements in a situation, predict future states, and identify which actions are appropriate in this situation. Mental models and schemata are based on experience [79]. Similar terms have been proposed by different theories in communication research. For instance, similar terms are *recipe knowledge* [26, p.57] and *habitualized actions* [26, p.71] in social constructionism (the sociocultural tradition in communication theory). In action-assembly theory, a related concept is *procedural knowledge*, which

been used to indicate people's estimates of the knowledge, abilities, role and goals of the robot [152]. We also link the concept of mental models to the concept of the *design metaphor*. This concept was discussed in the context of HRI by Deng et al. [65], and refers to how the associations that a particular robot design provokes lead people to have certain expectations regarding the way it functions. For instance, if a robot has a humanlike appearance, it appeals to a human design metaphor.

Mental models or design metaphors can also be provoked by means of the robot's behaviour. For instance, Cha et al. [41] report that when a robot has conversational speech abilities, people perceive the robot to have a higher level of physical capabilities than if the robot has functional speech, although this depends on whether the robot is successful at achieving its task. They conclude that functional speech is more effective at setting expectations at an accurate level. Conversational speech, in their experiment, consisted of phatic expressions, while functional speech concerned status information and next actions. This would suggest that conversational speech evokes expectations of social agency, while functional speech helps set expectations more correctly, as it is more in line with a device mental model or device design metaphor.

Thus, appealing to specific design metaphors or mental models may help people build correct expectations of how the robot functions. We do note that designers should take care not to reinforce societal stereotypes (e.g. regarding gender) when choosing to appeal to, for instance, human design metaphors.

Design Recommendation 3 Use an appropriate design metaphor to set human expectations of the robot and the interaction at a more accurate level.

2.8.4 Transparency

Another design recommendation is to make use of transparency mechanisms. Transparency involves communicating such things as the robot state, (accuracy of) sensing capabilities and currently active processes within the robot. By communicating the robot's limitations and internal processes, humans can form a more helpful mental model or perform behaviours that accommodate the robot's limitations [211]. Fischer's notion of transparency involves communicating reasons for robot failure, the robot's reliability, and robot awareness of the human to the human interaction partner [88]. The concept is related to Johnson et al.'s notion of *observability* [149], as well as *understandability*: making sure the robot's behaviour is understandable to its human interaction partner by making such things as the robot's beliefs and goals [122]. Transparency through visualization has been investigated in the context of robotics [43][211][249]. Communicating speech recognition results is an example of transparency regarding the sensing capabilities of the robot. A less taxing alternative with respect to human attention may be to indicate the accuracy of its speech recognition. Indicating whether the robot is currently processing input by changing the colour of a subset of its LEDs is an example of transparency regarding internal robot processes.

helps an individual to determine what to say or do next (sociopsychological tradition of communication theory). In Goffman's frame analysis, we find the terms *strips* and *frames* [177] (sociocultural tradition). Goffman set forth the notion of *primary frameworks* that, he says, individuals use as *"schemata of interpretation"* [107, p.21].

Regarding the example in Section 2.7.5, the success or failure in sensing the human's actions can be indicated by mechanisms that enhance system transparency. For instance, (1) the robot can indicate that the human's speech was not understood by verbally informing the human or by providing speech recognition results on a screen. (2) The robot may also provide transparency with respect to common ground by indicating which objects were recognized (verbally, or on a screen).

Design Recommendation 4 Use system transparency to communicate status information, sensing capabilities and currently active processes.

We note that finding the right level of transparency is a non-trivial task. A variety of communication modalities can be used to different effects, and other factors such as cognitive overload may start to play a role when a great deal of information is communicated to an end user. Testing the effects of different ways of conveying information as well as different types of information can be a labour-intensive process.

2.8.5 External influence

One difference between humans and robots is that external influences on the human side require interpretation by the human to have an effect on the human (with the exception of direct physical impacts), while external influence on the robot side is always direct. Humans can be directly influenced by impacting which information reaches the human or manipulating their body (e.g. turning their head to face in a certain direction). On the other hand, the robot has an 'open' nature; it is permeable to external influences (provided it is reprogrammable and reconfigurable), see also **Figure 2.2**. External influences and connections are not always problematic, but a few cases require further consideration. External influences and external data processing should be communicated to end users, especially in cases in which the end user's privacy is impacted. The external influences should be made explicit for the end user, and the end user should be asked for consent regarding external data processing. One can also think of the case of software updates. If it is important for the robot to maintain functionality even in the absence of a stable internet connection, the system developers should build a version of the system that still provides the desired functionality without external processing.

Design Recommendation 5 Communicate external influences to end users and, if possible and necessary, supply a product that is still functional without external processing.

Again, this problem is non-trivial, as regulations such as the GDPR need to be taken into account, as well as rights such as the right to privacy. The European GDPR regulation requires companies to provide end users with intelligible explanations regarding the way their data is used [80], which is also related to the issue of transparency as described in Section 2.8.4. Felzmann et al. provide considerations regarding robots, the GDPR and transparency, and propose a procedural checklist for implementing transparency within robotics development. The checklist includes steps such as identifying obligations, as well as stakeholders and their needs [84]. We also note that privacy and data processing must be even more carefully considered when it comes to social robots operating in public space. While guidelines for video surveillance with static cameras have been developed in the EU [76], for instance, social robots that move around in space autonomously and are equipped with cameras would require more specific guidelines and regulations.

2.8.6 Common ground

Achieving mutual predictability would require that the robot shares representations with a human interaction partner and 'understands' them in a similar way. Beliefs held by both entities can be considered common ground, although these beliefs are present in different ways in the human and the robot agent. For instance, on the robot side, beliefs can be stored in the form of logical statements. Note that with respect to common ground in the example in Section 2.7.5, *"cup"*, *'cup'*, **[object1]** and the actual cup are all different things.

With respect to instrumental uncertainty, common ground uncertainty and motivational uncertainty, interaction designers can choose to design robot behaviour in a way that reduces uncertainty of a human interaction partner. With respect to the example in Section 2.7.5, the robot could verbally communicate its goal at the start of the interaction to reduce motivational uncertainty.

Design Recommendation 6 Integrate specific robot behaviours to reduce instrumental, motivational and common ground uncertainty.

Solutions for disconnects in common ground include making robot capabilities explicit, e.g. by verbally or textually informing a user (system transparency). Another mechanism is to include external representations of the joint activity. This has been discussed by Clark, who gives the example of the chess board, a device that keeps track of the joint activity, that is, chess [48, p.45]. In HRI, screens and user interfaces can play the role of such an external representation. In the manufacturing domain, collaborative robot systems often include a graphical user interface that displays status information and task progress.

Design Recommendation 7 Use external representations of the joint activity to keep track of the accumulated common ground, if necessary.

Note that the 'common ground' indicated in the model will always be minimal compared to the common ground shared by humans. For instance, even if two humans cannot speak each other's language, they can oftentimes still communicate and understand each other. If a person encounters a robot that cannot interpret the language (or way of interacting) the human uses, the interaction will completely fail. In interaction between humans, we can assume a substantial common ground, which cannot be assumed in human-robot interaction [51].

2.8.7 Recommendations for modelling human-robot interactions

Based on the models surveyed in this paper and the process of modelling that led to the models proposed in this paper, we would like to give HRI researchers some recommendations with regards to modelling communication and interaction processes in HRI.

Model Design Recommendation 1 Define the level of analysis when discussing and modelling human-robot interactions or communication between humans and robots. It is not possible to include every level of discussion, nor every relevant factor, when modelling an interaction.

Model Design Recommendation 2 Clarify design choices and consider the assumptions made in choosing to model the interaction in a certain way. Be specific in presenting the supposed functioning of the human, robot and interaction. Make it explicit and keep in mind that there is a large variety of human bodies, abilities, behaviours and identities.

2.9 Limitations of the present work

There are limits to applying models of communication between humans to communication between humans and robots. However, we posit that models of communication between humans are a useful starting point, as they allows us to directly compare similar processes in communication between humans to communication between humans and robots, and humans likely bring expectations from communication between humans to their interactions with robots. This can give us insight into when, why and how communication failures may arise.

The model by Kincaid is not the only model of human perception, cognition and action (see e.g. [174]). Human cognition can most likely be depicted in a more accurate way, but this was not the aim of the current paper. The model by Kincaid was chosen as it meshes well with a joint action perspective, which is useful in the context of HRI and HRC. We hope we have given sufficient background to demonstrate that other types of models and research on communication theory can also be applied. We have outlined connections to different levels of discussion in communication theory. Here, we see another potential area of future research: models on the level of group communication and interaction, as well as on the level of organizations and society. At the group (or even media) level, one-to-many and many-to-many types of interactions can be considered.

Timing is of high importance in joint actions. The models in Figures 2.1 and 2.2 are depicted as processes, but do not detail exactly *when* communication is necessary. Changes in timing have an effect on how the action is interpreted. We have taken some initial steps towards incorporating the time dimension in the example in Section 2.7.5, but additional models and frameworks may be necessary. The model by Hellström and Bensch proposes that communication is necessary when one agent (agent X) determines there is a mismatch between the other agent Y's *estimation of agent X's state* and *agent X's actual state* [122]. This can be derived for the example in Section 2.7.5 in a similar way. However, the time dimension is of such importance that it deserves more prominence in a model of interaction. We would therefore also encourage other researchers to explore alternatives and come up with proposals that better express embodied, spatio-temporal and contextual aspects.

One question that arose during this work was the question of whether we can speak of communication at all when it comes to HRI, as using intentionalist vocabulary to describe robot behaviour may be too suggestive of human-level capabilities. However, as technologies advance, people may attribute communication capabilities to the robotic system anyway. This means that there are ethical consequences associated with this question, for instance regarding deception [53].

In this article, we connect the literature on HRI to (mainly) the broad field of communication theory. Our work has been especially influenced by symbolic interactionism and social constructionism [26]. Other influences are cognitive psychology and psycholinguistics (joint action), and theory on situation awareness [79]. We are aware that our thinking is highly influenced by tendencies common in European thought and research traditions (rational, focused on intentionality, cognitive, individual). For instance, it can be observed that the proposed model and the design recommendations place a large emphasis on cognitive processes and understanding. We invite other researchers to criticize perceived gaps in the arguments presented here and to propose alternatives.

2.10 Conclusion

The first aim of this article was to connect the research field of HRI to that of communication theory. We surveyed models of interpersonal communication from communication theory and focused our discussion on the transmission model of communication and transactional models of communication. We discussed communication and interaction models that are presently applied in HRI. We identified several models that fit a control paradigm of human-robot-interactions, and models that fit a social interaction paradigm. We identified and discussed several problematic aspects of existing communication and interaction models in HRI. The main problem we identified, is that often, the human and the robot are depicted as similar entities, while they clearly are dissimilar at the moment. This was in line with our second aim: to identify the asymmetries in human-robot interaction and communication. Differences in capabilities do not have to be problematic, as the robot's capabilities can be complementary to those of a human. However, communication failures as a result of these differences may arise. Another problem is that the interaction itself is often depicted in a simplified way, and understood as the 'sending of signals'. A joint action approach is more appropriate. The third aim of this article was to formalize an asymmetric model of joint action for HRI. We proposed the Asymmetric MODel of ALterity in Human-Robot Interaction (AMODAL-HRI). We did not aim to make the model as asymmetric as possible; instead, we aimed for the model to have similar processes on the human and the robot side to allow for direct comparison. This allows for identifying differences in a productive way: it allows for identifying asymmetries between human and robot capabilities and for proposing strategies to improve the robot's usability with respect to said asymmetries. In terms of practical applications, the model can be adapted to fit a specific technical setup. We demonstrated how the general model can be useful in practice, namely by means of the use of scripts as in the example in Section 2.7.5 and by comparing components and critically discussing the results of the comparison and differences with interpersonal interaction.

The main contribution of this work regards improving human mental models of robots, by investigating how interaction design can contribute to improving people's mental models of robots and their capabilities, in order to achieve successful human-robot interactions. By using this concept, we assume that people's previous experiences with technologies, objects, and even humans impact their expectations of interactions with devices such as robots. People's expectations can change by learning about or repeatedly interacting with the technology. We assume that if we achieve a better match between expected and actual robot behaviour, we will foster social acceptance and trust. Supporting accurate understanding of systems will help people know how to use the technology for their own goals, and help people rely on technology appropriately [171].

2.11 Future Work

As mentioned earlier, some aspects of the proposed model deserve more attention, such as timing in interaction, aspects relating to the environment (such as embodiment, physical space), and the involvement of other actors and team or group coordination. Future work can include surveying coordination frameworks and cognitive architectures for coordination, as well as approaches that model timing in interaction, with the aim of proposing additional communication and interaction models. We may also propose additional models that operate on different levels of communication (e.g. the level of group interaction, of organizations, the media, and society). Another interesting approach would be to adapt the model so that it integrates an existing cognitive architecture, for instance one based on a three-tiered architecture. This can be useful to see if the model still applies or breaks down. Finally, we propose that more work is required to detail how we can design transparent user interfaces for HRI applications.

2.12 Acknowledgements

This work was supported by Doctoral College TrustRobots, Technische Universität Wien. We would like to thank all anonymous reviewers, Michael Koller, and Guglielmo Papagni for their feedback on previous versions of this paper.

2.13 Conflict of interest

The authors declare that they have no conflict of interest.

CHAPTER 3

Co-design of Robotic Technology with Care Home Residents and Care Workers

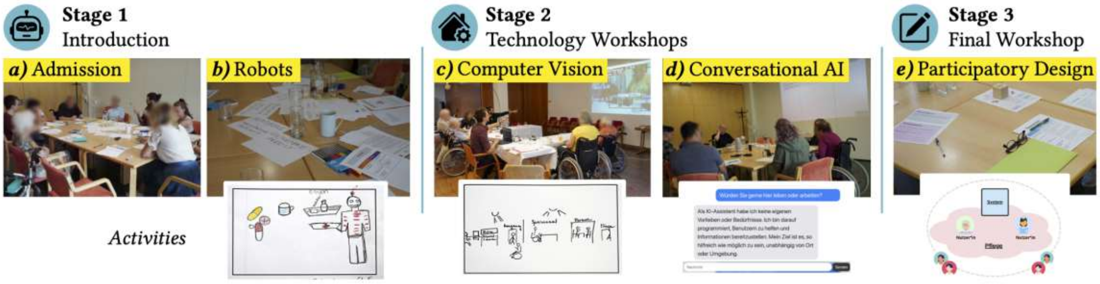
This dissertation chapter has previously been published as the following publication: Frijns, H.A., Vetter, R., Hirschmanner, M., Grabler, R., Vogel, L., Koeszegi, S.T. (2024) Co-design of Robotic Technology with Care Home Residents and Care Workers. In The PErvasive Technologies Related to Assistive Environments (PETRA) conference (PETRA'24), June 26–28, 2024, Crete, Greece. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3652037.3652070

3.1 Abstract

This paper reports on a co-design workshop series with residents and care workers in a care home, in which we ideate robotic technologies by starting from basic functionalities. We investigate whether introducing technology components to older adults and care workers in care homes enables them to imagine usage scenarios for (robotic) technology in care, and compared outcomes and engagement across groups: one group of care workers, one group of residents, and a mixed group. Having an interactive prototype prompted most response in the resident-only group. Inclusion of both care workers and residents highlighted the different and sometimes conflicting interests of the institutional context, contrasting responsibility for safety and the experience of living in a care home.

3.2 Introduction

Increasingly, Human-Robot Interaction (HRI) researchers apply participatory design (PD) or co-design (collaborative design [204]) methods to develop robot applications that are appropriate for



Würden Sie gerne hier leben oder arbeiten?

Als KI-Assistent habe ich keine eigenen Vorlieben oder Bedürfnisse. Ich bin darauf programmiert, Benutzern zu helfen und Informationen bereitzustellen. Mein Ziel ist es, so hilfreich wie möglich zu sein, unabhängig von Ort oder Umgebung.

Nachricht

Senden

Figure 3.1: Co-design workshops with residents and care workers in care homes, in three stages. Stage 1: Introduction with (a) focus group on Admission and (b) workshop on Robots and their strengths and weaknesses with a drawing activity. Stage 2: Technology Workshops on (c) Computer Vision with a drawing activity and (d) Conversational AI with interaction with a conversational prototype; Stage 3: Final Workshop on (e) PD with further development of scenarios using canvases.

people's needs in the care context and that make for more equitable design processes [204, 233, 170, 276, 135, 102]. An aging population and shortage of care workers pose societal challenges [166]. Though many applications for robotic platforms have been developed that aim to support care practices, by performing functions such as monitoring older adults, or providing cognitive or physical assistance to older adults, robot use is not widespread [166]. In studies on application of social robots at care homes, concerns have been raised regarding additional workload for care workers [187], fears that care workers will be replaced by robots [39, 166], the use of a robot causing care workers additional stress [40] and accessibility issues for older adults when (social) robots were used [39, 187]. Further, there are the issues of high cost, limitations to the maturity and readiness of existing robotic systems, concerns related to liability, privacy, technology acceptance, and limited research on long-term HRI [166]. Issues regarding accessibility, technology acceptance, and workload concerns make it all the more relevant to consider co-design methods for designing robot technologies.

In more general user-centered design methods, users tend to only evaluate (social) robot designs that are predetermined by researchers and developers [170]; such methods offer a limited space for users to be involved in the design process, and they are often treated rather as informants [204] than equal partners in the design process. Advantages of PD or co-design methods are improved representation of user needs and preferences in robotic technology design [233] and giving end users an enhanced sense of ownership [102]. Participatory design or co-design is an approach to design that enables stakeholders to impact design (in this case of robotic technology). Co-design processes can include using a variety of (creative) methods to elicit tacit knowledge, and includes communication and mutual learning between technology designers and stakeholders [135, 233, 102]. Tacit knowledge refers to experience-based knowledge that stakeholders have in relation to the context that they operate in. For example, care workers in a care home have a wealth of experience-based knowledge on how to care for, and communicate with care home residents. Co-design methods involving stakeholders and researchers create the opportunity for mutual learning. Researchers can gain a better understanding of what is important to end users and what the values of those end users are [135]. Stakeholders on the other hand can learn about technical capabilities from researchers [170]. This is also intended to address power imbalances between technology developers and users [170]. Communication about technology may decrease the information gap that exists on the side of care workers, so they are better able to imagine the use of the technology for their own purposes [40].

We report on a co-design workshop series on robots and technology components. The workshop series aimed to engage stakeholders in a care home, that is, care home residents and care workers, before a specific application scenario is chosen. The technical workshop topics in the series were robots, conversational AI, and computer vision. We aimed to present technical knowledge in accessible ways. For example, the session on conversational AI included a prototype that participants could ask questions with a chat interface. Moreover, we included creative activities to facilitate participation in verbal and tangible forms. We explicitly embedded the opportunity to iterate on ideas in a final workshop on PD. Bratteteig and Wagner [32] focus on design decisions as a key feature of user engagement in PD. Design choices is what participants in PD are involved in: the creation, selection, concretization and evaluation of such choices. In the research project Caring Robots // Robotic Care [98], we are in the phase of creation of choices, for which it is important not to close the space of potential choices too soon. For this sharing of power, we consider it especially important to enable stakeholders to learn about technical possibilities. The workshops took place with three different groups: one group of care home residents, one group of care workers who work at the care home, and one mixed group. Our main research question is: *How can we conduct participatory workshops about robotic technology components in ways that enable care home residents and care workers to envision use cases for technology in their care?* Sec. 3.3 describes related work. Sec. 3.4 includes an overview of the workshop series (see also Fig. 3.1). In Sec. 3.5 we report on the outcomes of the first four workshops, and in Sec. 3.6 on the outcomes and analysis of the final workshop. The contribution of this paper consists of the design of the workshop series, as well as recommendations for co-design workshops with care home residents and care workers in different group constellations (Sec. 3.7).

3.3 Related work

Examples of works that report on PD/co-design work of robotic systems with older adults, include e.g., [14, 170, 204, 102]. Common steps in PD processes are initial interviews, presentation of existing robots (for example by means of videos), developing a robot design, and presentation of ideas [233]. Antony et al. [14] report on co-design workshops with older adults to design robots that assist with physical activity. The process included initial interviews with nine older adults, a workshop with physical therapists, workshops with collaborative map making activities, personas, discussions, and sketching. Three students worked out designs that were critiqued in a final workshop with three older adults. Ostrowski et al. [204] describe a 12-month co-design process with older adults, concerning social robots for use in the home. Their process included initial interviews, art-based image making, living with the robot Jibo, discussing experiences, prototyping interactions with programming, generating social robot design guidelines, and reflection interviews. Lee et al. [170] and Randall et al. [221] report on a case study with older adults with depression. The process included initial interviews with older adults and therapists, a workshop in which existing robots were presented (videos and demonstrations), focus groups on designing a robot, a workshop on robotic parts, and a focus group with staff in which the designs by the older adults were discussed. The workshop on robotic parts included sensor cards that contained information about the sensors. Researchers gave demos of sensors (e.g., a camera). Participants were asked to choose sensor cards and talk about how these could support them in their life [221]. Gasteiger et al. [102] report on a four year project on the development of an assistive robot for mood stabilization and cognitive training applications for older adults that involved PD activities. Hsu et al. [135] report how they redesigned robot co-design workshops in order to make these more suitable for people living with dementia. The revised workshop series contained more familiar activities such as storytelling, singing by the robot, dressing up the robot and dancing with the robot [135]. Thunberg and Ziemke [276] report on a PD workshop with one group of care home residents, one group of older adults living at home and one group of care home staff. They asked participants what they know about robots and informed them about current robots. Moreover, they applied the PICTIVE method, which involved supplying participants with an image with four robots in a living room setting and materials for labelling the robots.

The main research gap addressed by our work regards the inclusion of both care workers and residents in the ideation phase of potential roles for robotic technology. Existing co-design processes for robotic technology with older adults tend not to include care workers in the ideation phase. Rather, care workers are consulted as informants or evaluators of designs by older adults, e.g., [14, 221]. Some other works also include care workers in the co-design process, but often they are not involved in the ideation phase that the older adults are part of (e.g. [221]). The one exception is [276], who ran a single workshop. Similar to [276], we include groups of care home residents and care workers in a care home. However, instead of one workshop, our process included multiple workshops and a variety of activities to engage participants in the co-design process. We chose to run the workshop in different constellations, including a group with both residents and care workers, to investigate how this influenced participation and idea generation. We expected different outcomes, as different stakeholder groups have different priorities and needs (see [169]). Therefore, it is worthwhile to investigate a longer co-design study with care workers and care home residents in the ideation phase. We ran 4-5 workshops with a duration of 2 hours each, with three different groups, involving different activities.

A second research gap is to enable learning about "robotic building blocks" through workshops on robotic components and software functionalities such as conversational AI. Co-design studies for robotic technology usually include existing robot embodiments and imply roles for robots [14, 204, 221, 276]. Existing works on participatory design for HRI often include existing robotic platforms (e.g., [204, 170]) and predetermined application scenarios (e.g. [204, 14]). Randall et al. [221] write that showing finished robots resulted in older adults either rejecting or accepting the system as a whole rather than considering modification. An advantage they report is that participants referred back to robot forms shown earlier [221], and it is useful to get responses to particular design instances. However, the question remains whether there are other roles for technology that meet needs that are under explored, but that can come to light when technology is introduced in a different way. Existing robots imply certain ways of functioning and certain possibilities for social interaction, e.g., the robot Pepper implies a role or relation of a humanlike conversation partner. Instead of implying certain roles for the technology, or introducing existing robot embodiments, we introduce technology components and investigate whether this leads to participants being able to imagine different roles for technology. With technology components, we mean software and hardware components that are commonly used in robotic systems for HRI scenarios and social robotics. Similar to [221, 170], we introduce technical components and discuss robots, though our work differs; we include two dedicated workshops on computer vision and conversational AI, and focus on the care home context. In our work, we aim to leave the robot's role open, and engage in ideation on the basis of an understanding of technical capabilities. We relate this to the notion of sharing power in decision-making in PD [32].

3.4 Workshops

We report on a series of co-design workshops at a care home, see Fig. 3.1 for an overview. Workshops were conducted with three different groups, one consisting of residents, one of care workers, and a mixed group of both. The topics of the workshops were an introductory discussion on admission, robotic systems, conversational AI, computer vision, and a final workshop on PD in which ideas that had come up in previous workshops were developed further. The first session on admission was intended to suggest a care practice/life period that a robotic technology may support, and to establish relationships. The second workshop on robotic systems aimed to illustrate different components that are part of robotic systems, to discuss common assumptions about robots, and strengths and weaknesses of robotic technology. We chose to introduce conversational AI functionality, as speech interaction is a common aspect of communication in HRI. Computer vision is an aspect of robotics that enables a robot to perceive and navigate its environment and to do meaningful tasks in relation to objects and people in the environment. Based on recommendations in the literature, we included opportunities for iteration in the workshop series, a variety of activities to promote different forms of engagement [204, 233], and convergent and divergent activities [204]. We considered limitations regarding sensorimotor skills and cognitive capacity in workshop material design, e.g., by having workshop facilitators draw for older adults if required [233]. We conducted the workshops at two care homes in Austria. These were care homes of a care organization that is a research partner in our project. Each workshop was run by 3 to 4 facilitators from our pool of researchers with expertise on PD, HRI and sociology, and one care expert. Some facilitators were kept constant between workshops to ensure familiarity for participants. The workshops took place in a separate space that was used for occupational activities, where participants gathered around a table. A projector was used for the Conversational AI and Computer Vision workshops. The workshop series was peer reviewed by the Research Ethics committee at our university and was conducted in line with ethical guidelines at our university.

3.4.1 Recruitment

We recruited participants with support of the care organization in two different care homes (care home A and care home B). We approached participants personally before the first workshop and discussed the study and their potential involvement. We prepared an information sheet about the workshop, discussed the informed consent and data consent forms with potential participants, and answered questions, after which they were asked to read and sign the documents. We aimed to include the same participants across all workshops. However, due to limited availability (e.g., care workers being on vacation; care recipients not feeling well), new participants could join at any stage except the last workshop.

ID	Gender	Age	Role
CW1	F	44	Dementia expert
CW2	F	35	Qualified nurse, care station lead
CW3	F	37	Quality manager
CW4	F	37	Care worker, quality manager
CW5	M	43	Care assistant
CW6	F	55	Qualified nurse
CW7	F	41	Qualified nurse
CW8	M	55	Qualified nurse, quality manager
CW9	M	33	Qualified nurse
CW10	F	36	Qualified nurse, care station lead
CW11	F	48	Social worker
CW12	M	43	Care home director
CW13	F	59	Social worker
R1	F	83	Care home resident
R2	F	91	Care home resident
R3	F	93	Care home resident
R4	M	59	Care home resident
R5	F	61	Care home resident
R6	M	91	Care home resident
R7	M	74	Care home resident
R8	F	80	Care home resident
R9	F	87	Care home resident
R10	F	91	Care home resident
R11	F	75	Care home resident
R12	F	82	Care home resident

Table 3.1: Overview of workshop participants.

Inclusion criteria for care home residents were that potential participants were able to participate and engage in 2-hour workshops, were expected to enjoy and benefit from it (based on the judgement of care workers involved in recruitment), and that they were unlikely to experience stress or other negative consequences from participation. Moreover, a requirement was that they were legally and practically able to give informed consent prior to the first workshop, as well as process consent (renewed consent for subsequent workshops). We recruited care workers of diverse qualification levels (i.e., care home lead, station lead, qualified nurse, nursing assistant), who took part within the context of their work and were reimbursed using project funds.

3.4.2 Participants

In total, 25 participants took part in the workshop series, see Table 3.1. Of the 25 participants, 13 were care workers (4 men, 9 women, age M=43.5 years, SD=8.4 years), and 12 were care home residents (3 men, 9 women, age M=80.6 years, SD=11.5 years).

3.4.3 Workshop series

Five different workshop formats were prepared. Each workshop was conducted three different times, one time with residents, once with care workers, and once with a mixed group. In the resident group, a care worker was present in case residents needed support. Due to short-term changes in the availability of participants, the introduction workshop with the group discussion on admission and the robotic assumptions workshop were combined for care home B. See Table 3.2 for an overview of the workshops per location and per group. The start of every workshop included a short introduction with a summary of the workshop contents, and the possibility to ask questions. New facilitators were introduced. A facilitator would mention that the workshop would be audio recorded and switch on the audio recording devices. The main components of every workshop included explanations of technologies, creative activities, and group discussions. Handouts were prepared for each workshop. All workshops and workshop materials were in German. The workshops took place approximately once every 1-2 weeks. Each workshop took two hours. Note that the order of the Computer Vision and the Conversational AI workshop are switched for care home B.

Introduction workshop and group discussion on admission

The aims of the first workshop were to get to know and build relationships with the participants, and to present the structure of the workshop series. The facilitators of the workshop series introduced themselves and the topics of future workshops, describing the technologies that would be introduced and the planned activities. Participants were asked to introduce themselves. After the general introduction, the remaining time was spent in a focus group format on admission. Residents were asked to reflect on their admittance to the care home, and care workers were asked about their experiences when new residents are admitted. Discussion themes included, for example, documentation of care actions by care workers in the central documentation system (WS1B-CW), instances in which the admission of residents was unplanned (WS1B-R), the acclimatization period when residents arrived in the care home (WS1A), and biography work (WS1A). Biography work refers to care workers collecting biographical information on residents so that care workers are better able to provide person-centered care that takes a residents' preferences and life history into account.

Robotic assumptions workshop

The focus of this workshop was on robots. First, the themes from the previous workshop were revisited. Participants were asked to consider how a robot could support them, and were provided with a worksheet on which the participant could visualize or textually describe the robot performing a certain task. If a participant needed support with drawing, they could instruct a facilitator regarding what should be drawn. Participants were asked to tell the group what they envisioned their robot to do. After the creative activity, we discussed how robots work and what kinds of components robots require to function. We used a schematic depiction of a robot containing different robot components. The facilitators gave an example using the drawings from the creative activity. Then, the facilitator introduced aspects that robots are good at (e.g., availability and precision, processing speech and translation) and not good at or not capable of (e.g., empathy and emotion). These were inspired on the concept of *robotic superpowers*, which refers the idea of making use of what robots are particularly good at to develop alternative forms of social interaction rather than building robots that have

Workshop (WS)	Topic	Care home	Group	Participants (codes)	Facilitator
WS1A	Admission	A	Mixed	CW1, CW2, R1, R2, R3, R4	F1, F2, F3
WS2A	Robots	A	Mixed	CW1, CW2, CW3, R1, R2, R4	F1, F2, F4, F5
WS3A	Computer Vision	A	Mixed	CW1, CW2, R1, R4	F1, F3, F5
WS4A	Conversational AI	A	Mixed	CW3, CW12, R1, R4	F1, F3, F4, F6
WS5A	Participatory design	A	Mixed	CW2, CW12, R1, R4	F1, F2, F3, F5
WS1B-R	Admission and Robots	B	Residents	R5, R6, R7, R8, R9, R10, CW11	F1, F2, F4
WS2B-R	Conversational AI	B	Residents	R5, R6, R7, R8, R9, R10, CW13	F1, F2, F3, F5
WS3B-R	Computer Vision	B	Residents	R5, R7, R10, R11, R12, CW11	F1, F2, F5
WS4B-R	Participatory design	B	Residents	R5, R7, R8, R10, R11, CW13	F1, F2, F5, F6
WS1B-CW	Admission and Robots	B	Care workers	CW4, CW5, CW6, CW7, CW8, CW9, CW10	F1, F2, F4
WS2B-CW	Conversational AI	B	Care workers	CW5, CW6, CW7, CW8	F1, F2, F3, F5
WS3B-CW	Computer Vision	B	Care workers	CW6, CW7, CW8, CW10	F1, F2, F5
WS4B-CW	Participatory design	B	Care workers	CW4, CW5, CW6, CW7, CW8, CW10	F1, F2, F5

Table 3.2: Overview of all the workshops. Facilitators are included in the table to demonstrate consistency across workshops. The facilitators presented content, asked the participants questions, and answered those by participants. For an overview of residents (R) and care workers (CW), see Table 3.1.

humanlike capabilities [197, 74, 7, 75, 299]. This was followed by a focus group (a discussion setting with participants to explore their attitudes [233]). For the focus group, questions were prepared regarding what such a technology would mean for care, for privacy, and what participants took away from the workshop (these questions were also asked at the end of the Computer Vision and Conversational AI workshops below). Depending on the energy of the group, the creative activity was done a second time.

Computer Vision workshop

This workshop started with a short demonstration of RGB and a depth camera. The live camera stream was shown as a 3D point cloud on the projector. Next, participants were asked to imagine how a technology could support them with seeing. This was a drawing/writing exercise. The workshop continued with an example using pictures of cats and pictures of animals with similar features (e.g., whiskers, pointy ears). This example was given to illustrate the difficulty of specifying an exhaustive set of rules to assign the correct category label to a detected instance. After a short break, the facilitator gave a high-level explanation of computer vision and neural networks with a step-by-step example of how to detect handwritten digits. Two demo applications were shown. One demo application was on object detection. In the demo, objects and people in the space were detected and classified. A visualization was shown of the camera stream, with bounding boxes drawn around the objects. Another demo application was on detecting body pose, with skeleton joint positions that were drawn on top of the camera stream when a body pose was detected. There was further discussion in a focus group setting.

Conversational AI workshop

For this workshop, a system was developed using OpenAI Chat Completion with the gpt-3.5-turbo model [201] and macOS text-to-speech and speech recognition. This was combined with a visual chat interface and a microphone and speaker. Participants could speak into the microphone to ask a question and the OpenAI API was used to generate a response. To prevent sending personal data, controls were built in so that the facilitator could check and alter the prompt before sending it to the OpenAI servers. Similarly, server responses were first checked by the facilitator before being shown to participants.

The workshop started with a demo in which the conversational program wished the participants a good morning. The researcher explained, among other things, that large language models use many sources of information and that these work by taking the next word in a sentence that has a high likelihood of being correct. Participants were given a worksheet with the question what they would ask a conversational program. They could pose these questions to the system. Then, the researcher explained what such conversational programs are good at (e.g., translating) and bad at (e.g., no personalization is included). There was a final discussion in a focus group setting.

Participatory Design workshop with scenario development

This workshop included a short introduction on the topic of PD and emphasized that participant input was valued in the context of the workshop series. We brought in scenarios that participants had come up with during previous workshops (from discussion themes, worksheets) and worked these out in smaller groups of 2-3 participants and 1-2 facilitators. The workshop was shaped differently for the resident group, in response to how they had been able to contribute and participate in previous workshops (see below). Moreover, different scenarios were selected for each group, depending on outcomes of the creative activities in prior workshops, as described in Sec. 3.5. The scenarios and the discussions in the smaller groups are described in Sec. 3.6.

Canvases were prepared, inspired on social robot co-design canvases by Axelsson et al. [16], which are intended as a tool to structure multidisciplinary design processes. For our workshops, we prepared individual, smaller canvases with one topic per canvas, which were mainly used to moderate the discussion. Participants were presented with one scenario (e.g., a system for documentation support for care workers) and several functionalities (e.g., making a summary of the night shift). They were asked to select functionalities to work out further in the discussion. Questions on the canvases focused on topics such as who would use the technology, the location and form of the technology, when the technology would be active, how the functionality is achieved at the moment, privacy, impact on work/life, input modalities, and output modalities. See Fig. 3.2 for the first canvas that presents an overview of the discussion topics. The workshop ended with a final discussion, a feedback round, and discussing images of robots, including pictures of robots at our lab.

3.4.4 Collected data

Collected data consisted of audio recordings of the 13 workshops, 13 audio recordings of reflection discussions by researchers directly after each workshop, and pictures of worksheets participants completed. The audio files of the workshops and the reflection discussions were transcribed with a local installation of OpenAI's Whisper [202], and were manually checked, corrected, and speaker changes were annotated.

3.5 Outcomes of the creative activities

In this section, we report mainly on the engagement with creative activities and creative activity outcomes from the workshops described in Sec. 3.4.3-3.4.3. An overview is provided of the types of questions that were asked in the Conversational AI workshop and the types of scenarios envisioned in the drawing activities.

3.5.1 Questions to a conversational program

Participants were asked to reflect on what they would like to ask a conversational program, and to write these questions down. Some questions were entered in the system through speech or typing in a demo setting. In W4A, questions were asked on practical matters such as the weather, who has the night shift, translation, menu and activities for the day, doctor's appointments, and what time it is. There were also questions that were intended to make sense of the system. One resident specified instructions (e.g., bring me lunch in my room). Other application ideas included input in the documentation and engaging in conversation with residents if a care worker is occupied elsewhere. Questions that were asked by residents in WS2B-R included questions about practical matters, such

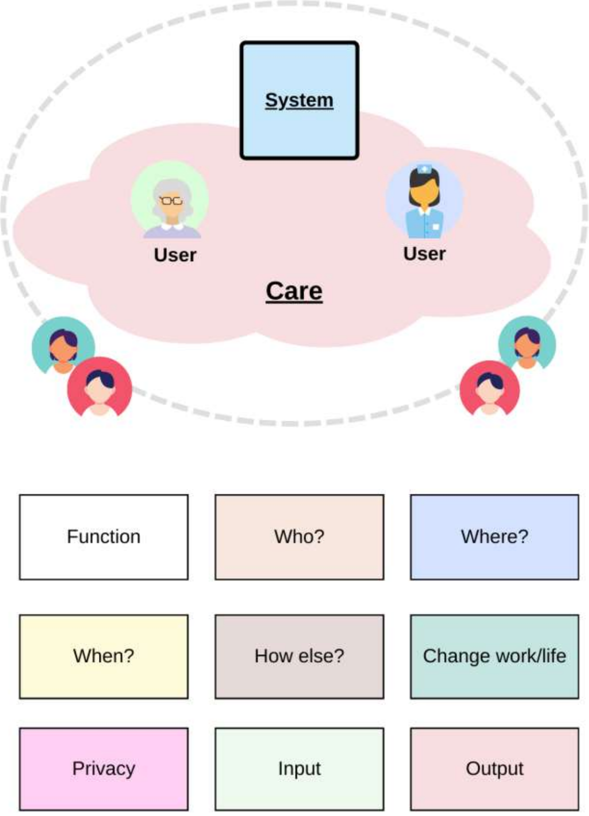


Figure 3.2: Main canvas that presents an overview of the topics of the other canvases for discussion during the final Participatory Design workshop.

as what to cook tomorrow or how to reach a relative, and personal situation questions (e.g., *"when will I be able to walk again?"*). Several questions related to making sense of the system, see 3.7.2. In WS2B-CW, there were practical questions about diagnoses, active agents in medication, contact data of doctors, prices of medical products, answering legal questions, and how the night shift went. There were questions regarding capabilities of the system, for instance whether it recognizes if a person is sad.

3.5.2 Drawing activities

Both the *Robots* and *Computer Vision* workshops contained activities where participants were asked to envision a scenario by drawing and/or writing on a sheet. Drawings were analyzed by combining the verbal descriptions of drawings from the workshop transcripts and the worksheets, which were qualitatively analyzed and coded regarding the application scenario that the participant described, see also [204].

Overall, ideas by residents from the *Robots* and *Computer Vision workshops* included robots for physical assistance and rehabilitation, bringing objects/breakfast to residents, reading small fonts, and providing information such as the plan for the day. The main themes that were suggested by care workers related to some form of documentation support, including documentation in all languages, automatically doing measurements and entering these in the documentation, inferring care plans, and providing training for care workers. Secondly, several ideas related to pressure injury prevention, classifying wounds and tracking wound healing. There were ideas relating to fall prevention, tracking the locations of residents who have a tendency of walking away, and providing information to relatives and residents. Ideas that were suggested once included robots for distribution of medication or food and for physical activation of residents.

Regarding engagement of participants during the workshops, in the *Robots* workshop all participants in W2A came up with ideas in the first and second round. For two of the residents, the concept of the robot failing or having weaknesses from the facilitators' explanations seemed to have made an impression. During the second creative activity, one resident talked about a robot that fails while another robot comes to help the failing robot. Another brought in the concept of learning, suggesting a scene in which a robot had to learn to cook different recipes. In WS1B-R, the drawing activity was met with skepticism. For example, R7 stated: *"I return this empty. What can happen? I'm lying on the floor and a robot comes and helps me out. I don't believe it."* Two out of six residents drew or wrote something. In WS1B-CW, all care workers completed a worksheet. One person drew a human and a robot and stated not being able to imagine the robot. One participant drew two robots, stating that they should *"communicate at eye level"*, which means that there is no hierarchy. In the *Computer Vision* workshops, all participants completed a worksheet in W3A. R1 first stated not having any idea. After being prompted by R4, who asked if the robot could help her read small font, R1 proposed a robot that read out text messages she gets on her phone. In WS3B-R, three out of six wrote or drew something on the worksheet: a reading aid, a nightstand with a lamp and radio, and one resident could not imagine anything and drew question marks. In WS3B-CW, all four care workers came up with ideas.

3.6 Analysis of the PD workshop

The last workshop is described in more detail here, as it is exemplary of the iteration step we embedded in the workshop series to work out participant ideas further. At this point, participants had received all information and were familiar with us and the group, which makes it most relevant to see how our approach impacted the outcomes and engagement in different groups. For the last workshop, the ideas that were proposed by participants were iterated on in smaller groups. Selected application scenarios were further discussed in the last workshop (Sec. 3.4.3). The transcripts of the six group discussions in WS5A, WS4B-R and WS4B-CW were analyzed using summative qualitative content analysis [186]. This approach was used to reduce the material to obtain first results on the focal points of the contents of the scenario discussion, the group dynamics, and the ways of relating to the topic of discussion. The researchers used MAXQDA to categorize and translate the material into summaries for each group discussion. F2 and F4 jointly discussed these summaries of categorization by checking back with the original material for additional validation [186, p.165]. This process was based on the idea of Kuckartz [164] who developed Mayring's method further by taking into account case comparisons within qualitative content analysis. Being able to compare the cases (the different group constellations) was crucial to answering our research question regarding participation and ideas of the different care stakeholders who were involved.

3.6.1 Mixed group

In WS5A, five scenarios were brought in based on ideas from prior workshops. The five scenarios were (1) a personal assistant for residents that can bring objects and provide information, (2) a monitoring system, (3) a documentation system for care workers, (4) a digital companion for biography work, and (5) a system for rehabilitation and fitness support. Participants were divided into two groups based on their interests.

In one of the groups (CW12, R1, F2, F3), the documentation scenario and an assistive robot scenario that could perform tasks such as bringing food were discussed. Speech input, translation capabilities, and automatic entry of data from different sensors (e.g., blood pressure) were deemed beneficial for the documentation system. The conversation was carried by the care worker. The resident opted to be in this group out of curiosity (and stated not needing technology that brings her something or informs her of something). The resident related part of what the care worker said to her own experience, for example to illustrate that she sees the use of a translation functionality or that an emergency system in case of falls should be 100% reliable.

The discussion in the other group (CW2, R4, F1, F5) focused on a potential monitoring system that could measure vital signs of care recipients. A hypothetical system of wearable sensors was discussed that would alert care workers if values were out of the norm. Preferences were discussed, such as manually setting personalized thresholds, between which the system should not alert care workers. Both the care worker and the resident talked about current care practices and how a system could function, look, etc. R4 was younger than the average care home resident and belonged to a different generation. He brought a strong focus on logistics and considered the care relation as a whole, not just from a care recipient's perspective, for instance suggesting giving care recipients access to the data and making care relations more equal.

3.6.2 Residents group

In WS4B-R, participants had suggested few potential application scenarios themselves in prior workshops. However, the topics of biography work and feelings of being monitored had previously come up. Therefore, the scenarios of a digital companion for biography work and a monitoring system from W5A were brought in. For this group, short stories were prepared that gave more detail on particular aspects of the scenario.

One group (R5, R8, R10, F1, F6) discussed a digital companion for biography work. First, residents were asked how biography talks were conducted. These talks happened during or shortly after admission to the care home. Residents then talked about their own biography, though this was not what facilitators intended talking about. Talking about their own biography was dominant in this group. F1 introduced a scenario on biography talks between a resident and a care worker, which involved a hypothetical robot that asked a question. The residents rejected the idea of talking directly to a robot, which they deemed "impersonal".

The other group (F2, F5, R7, R11, CW13) discussed the topic of monitoring. This mainly led to one of the residents indicating she is wearing a watch that notifies care personnel when leaving the ward. The resident had voluntarily agreed to wearing the watch, but expressed feeling monitored. This resident thus shared personal experiences relating to technology use and feeling monitored.

3.6.3 Care workers group

In WS4B-CW, only the documentation scenario was brought in, as participants had often mentioned and proposed functionalities relating to this scenario in previous workshops.

In one group (CW5, CW8, CW10, F1), preferences were discussed relating to the design of a care documentation assistant. The functionality of supporting translation and documentation in different languages came up repeatedly. Hygiene and safety aspects relating to portable devices were discussed, as well as privacy concerns relating to audio recording. Care workers reported on the advantages and disadvantages of specific use cases, and practical considerations regarding form factors. They discussed mostly in relation to how a technology would impact their workflow (e.g., disadvantages such as technology resulting in too much documentation, worries about people assuming the technology is always right). One care worker suggested a methodological change: an additional topic for a canvas, namely a sheet on risks, negative effects, unwanted consequences.

In the other group (CW4, CW6, CW7, F2, F5), care workers discussed potential modification and use cases of a documentation system intensively, going through the topics of the canvases, advantages and disadvantages. The care worker with the lowest qualification level spoke least.

3.7 Co-Design with Care Workers and Residents

In this section, we outline key takeaways from the workshop series for co-design processes with care home residents and care workers.

3.7.1 Complementary perspectives of residents and care workers

Takeaway: Including both care workers and residents in the co-design process gives a more comprehensive picture of the care practice, e.g., care workers may often focus on safety aspects while residents can give personal accounts of the way they experience this focus on safety.

Competing values and expectations need to be negotiated in shared settings that take perspectives of both care workers and residents into account. For example, we contrast a residents' experience of feeling monitored with the positive attitude of care workers to monitoring applications we observed, e.g., in creative activity results.

What can be observed in care worker ideas is a focus on workflow, procedures, and resident health and safety. Care workers carry responsibility for safety of residents. Documentation is a way to document care actions that have taken place, to provide legal evidence, keep track of developments over time, and make care plans. CW12 envisioned in W5A that an improved documentation system with speech input could lead to higher efficiency and could lead to improved security when combined with monitoring systems. In WS4B-CW, care workers participated in the discussion from the perspective how the technology would impact their workflow.

Residents on the other hand bring in accessibility perspectives and can provide personal accounts of the way they experience (existing) technologies such as monitoring. In the story of the resident in WS4B-R, Sec. 3.6.2, a negotiation was described between the care staff and the resident. In an institutional context, there can be pre-existing issues that put a large onus on safety, but this can result in feelings of surveillance and technological solutions being experienced as a compromise. Different stakeholders can have conflicting interests and views (compare [169]). One accessibility issue occurred in W4A, when a resident could not hear the tone that indicated that the system was listening. R7 brought up accessibility perspectives in WS2B-R in relation to the conversational system, stating that many residents had difficulties hearing and understanding, and that they would not be able to interpret or use such a system.

3.7.2 Interpreting technology together

Takeaway: An interactive conversational prototype facilitated the workshops as it took a role in the group discussion and joint sensemaking of the technology.

Having a concrete prototype that participants could interact with and ask questions helped to get them engaged. In the residents group, this activity provoked most engagement compared to other creative activities. While for workshop WS1B-R and WS3B-R there was little response to the drawing activities, the possibility to ask the conversational system questions in WS2B-R was responded to much more. Besides residents interacting with the technology itself, it also led to interaction among residents. For example, after one resident asked a question about a singer, this prompted further questions about composers and discussions about favorite operettas among the residents.

The prototype also enabled learning; R9 asked *"What is a conversation program?"*, reporting that she could not imagine anything. The response of the conversational system was: *"A conversational program is a software or computer program designed to have conversations with people. It can answer questions, provide information, or conduct general conversations. It uses artificial intelligence to generate human-like responses."* R9: *"Aha, yes, now I understood it."* Resident R7 stated doubting that everyone had understood it. The care worker explained: *"That is a computer program, (...) that can answer our questions."* R9: *"Yes, it is called a conversation program."* The facilitator let the system repeat the answer in a simplified way. R9 then asked *"But it can only answer, what you ask it, I believe?"*

This illustrates how the participant tried to make sense of the system, and that this was a joint process of residents, the care worker, facilitators, and the system. On the other hand, questions in the care worker group related to difficulties in the care practice, such as legal guardianship legislation (see Sec 4.1), and integration of such technologies in care practices.

3.7.3 Openness and Ideation

Takeaway: Openness regarding the application will mean that participants may not be able to imagine anything or may go in the direction of what they already know in terms of care

practices and existing technologies they use. This is helpful for finding out their needs in a co-design process, but may be less so if the aim is out-of-the-box thinking.

Care workers appeared to change their interpretation of technology over the course of the workshop series, in order to use the openness of the activities to consider applications that supported them. In the feedback round in WS4B-CW, CW7 stated: *"At the beginning I was a bit skeptical, because I think just like the others, could not imagine, now a robot in care, but now that we've seen more of it, these are technologies that can possibly help us, or can support us, I now have so many ideas."* CW4 in WS4B-CW stated that for her, a rethinking process had taken place. For her, it became possible to focus on supporting processes in care with a networked way of thinking, instead of keeping a focus on the device itself. In some cases, this resulted in shifting away from robots and towards more generally considering how technology may be useful in care. This is exemplified by CW1, who stated in W2A: *"I'm not a fan of robots, but I was thinking what could be useful for us in care."* when describing the result of the drawing activity. Care workers suggested applications that go more in the direction of general health IT (e.g., general care information systems, skin management [17]). This could also be due to the practices and technologies that they are already familiar with and where they see potential for improvement.

Few residents in the resident-only group proposed ideas in the drawing activities. This could partially be due to the openness regarding the application scenario. In WS4B-R, responses often went in the direction of their own lives/biographical information, either because they prefer talking about this or because they have a hard time envisioning the use case but want to respond still. The discussion focused often on residents' biographies, which also means that there is a potential way to access their experience there. The openness of the application scenario may also have led to residents being unsure about the intentions of the technology, which could lead to skepticism or replacement concerns. In WS1B-R, R8 responded as follows to the drawing activity: R8: *"(...) our nurses, they do it very well, the body care. Starting from the face to the toe. I don't know if that's how it does it, the way a nurse does it. She does it more sensitively. I don't know what its attitude is. Excuse the expression, the tin idiot. (...) It does not have any feelings."*

The difficulty of imagining a technology that does not exist yet was more easily overcome in the care workers group, but residents needed more support. The aim was to leave the application scenario open to create space to imagine new roles for robots. It can be difficult for participants to deal with this openness. Hence, activities should be scaffolded appropriately. When working with residents, it is important to reserve time for and include activities that enable sharing of personal stories and relating to other group members, especially for larger groups. Working with smaller groups in which residents collaborate with care workers can support them in reflecting on technology. Focusing on technology in resident-only groups requires more active moderation. In the care workers group, technology demonstrations and mutual learning established rapport, made them aware of design opportunities afforded by technology functionalities, and increased their eagerness to contribute.

3.7.4 Group dynamics

Takeaway: Group activities offer the advantage of different perspectives on potential applications, but it is important to be aware of group dynamics. For instance, there may be profes-

sional hierarchies among care workers.

In the care worker group, care workers had different job roles, which also meant that there was potentially a hierarchy in the care worker group, which could also be reflected in how much each of them spoke. However, care workers with different roles have different perspectives on the potential applications (e.g., a station lead suggested a system for making shift plans in WS3B-CW), which can be a benefit. Particular group dynamics can make replication of co-design studies difficult [221]. Among the residents, we observed large differences regarding technology skills and their apparent ability to understand and relate to the contents of the discussions, and generational differences. Some residents were more familiar with technologies, and had a smartphone or laptop, while others were less familiar. Similarly, Antony et al. [14] report a large diversity in the older adult population. The engagement of residents in the mixed group was very different from the engagement of residents in the resident-only group. It is likely that group dynamics played a role there, with actively participating care workers supporting/encouraging resident participation. While R1 in the mixed group often says that she cannot imagine things, in this context she is with other people who propose ideas, and then R1 also suggested ideas. It can be beneficial to consider how to make the environment encouraging for idea generation, even if the participant may not be immediately comfortable with ideation or creative activities.

As a researcher, it is important to familiarize oneself with the qualifications, responsibilities and hierarchies that exist in the care context under study, and to balance inclusion of care workers at different qualification levels of relevance. To balance group and power dynamics, designers can use active moderation to include participants who speak less, smaller groups where participants are at similar hierarchy levels, and take time to establish a practice of sharing ideas.

3.8 Limitations and future work

Limitations included that residents who participated in our study were not representative of the large proportion of care home residents who are people living with dementia, for whom different participation formats will be necessary, see [135]. While we had initial individual talks on the workshop series with potential participants, it is highly advisable to have individual preparatory talks with potential participants regarding their technological skills, to be better able to prepare workshop materials that are suitable to individuals. This is also an opportunity for residents to already share biographical information that they may be happy to share. Combining the first two workshops for care home B may also have impacted the results, as this meant there was less time for relationship building compared to care home A. It is advisable to have smaller groups to be able to more actively involve all participants.

For future work, inclusion of younger participants or people who are cared for at home in the co-design process can be considered. Residents do have the perspective of what it is like to live in a care home. However, future generations are likely more technically skilled and may be more interested in technology being part of their life.

Future work for our project will include further developing the ideas that were worked out by participants. Documentation was one of the central themes among care workers and reported to take a lot of time. There may be large potential gains regarding quality of documentation, care, work experience, and efficiency there, but also open questions.

3.9 Conclusion

This paper reports on a co-design workshop series with three groups: one group of care home residents (4 workshops), one group of care workers (4 workshops), and one mixed group (5 workshops). Workshop topics were admission to the care home, robots, computer vision, and conversational AI. The final workshop included the opportunity to iterate on ideas: participants further discussed application scenarios that were suggested by them in previous workshops.

We found that care workers and care home residents bring in complementary perspectives, which can be conflicting at times. Care workers reflected on advantages and disadvantages in relation to their workflow, while residents bring in personal perspectives that come with living in a care home. The interactive prototype in the conversational AI workshop was integrated in the conversation between residents in the residents-only group, while care workers reflected on its integration in their work practices. Over the course of the workshop series, care workers started to consider how technologies could be useful in their work practice, while we observed several cases of residents remaining skeptical, especially in relation to the idea of robots in care. Regarding group dynamics, we remark that among care workers, professional hierarchies exist. While care workers with different roles can bring in a rich set of perspectives, it is important to remain aware of potential hierarchies. Moreover, the population of care home residents is diverse, with different levels of technology knowledge.

Our work emphasizes the importance of participatory formats for robots in care, as such formats give space for different, conflicting voices to be heard. Moreover, it highlights the challenges of openness in co-design and suggests ways participants can jointly interpret technologies in such settings.

Acknowledgements

We would like to thank all participants in this study. We thank Muhammad Saleemi for his support with transcriptions. This work was funded by a #ConnectingMinds grant to the project Caring Robots // Robotic Care (CM 100-N). The research proposal was peer reviewed by the ethics committee at TU Wien, under case number 035_12052023_TUWREC.

CHAPTER 4

Design Guidelines for Collaborative Industrial Robot User Interfaces

This dissertation chapter has previously been published as the following publication: Frijns H.A., Schmidbauer C. (2021) Design Guidelines for Collaborative Industrial Robot User Interfaces. In: Ardito C. et al. (eds.) Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12934. Springer, Cham. https://doi.org/10.1007/978-3-030-85613-7_28

4.1 Abstract

Collaborative industrial robot (cobot) systems are deployed to automate tasks or as a tool for Human-Robot Interaction (HRI) scenarios, especially for manufacturing applications. A large number of manufacturers of this technology have entered the cobot market in recent years. Manufacturers intend to offer easy control possibilities to make cobots suitable for different user groups, but there are few evaluation tools for assessing user interface (UI) design specifically for cobots. Therefore, we propose a set of design guidelines for cobots based on existing literature on heuristics and cobot UI design. The guidelines were further developed on the basis of modified heuristic evaluations by researchers with robotics expertise, as well as interviews with cobot UI/User Experience (UX) design experts. The resulting design guidelines are intended for identification of usability problems during heuristic evaluation of the UI design of cobot systems.

4.2 Introduction

This paper is concerned with the design of robots and associated UIs that are part of collaborative industrial robot systems (cobots). Cobots are systems that are intended for collaborative operation, i.e. the concurrent execution of tasks by human(s) and robot(s) in the collaborative workspace (as defined in [143]). Such a system is not collaborative by itself; rather, the collaborative nature arises from the way the application and interaction are designed. A cobot system usually includes a robotic arm and a graphical user interface (GUI) with which, for instance, a factory worker can program the robot to do a specific task in a manufacturing context. The last decade has seen increased interest in the introduction of cobots to the shop floor [77]. Various manufacturers of automation solutions (e.g., ABB, COMAU, FANUC, KUKA, Stäubli, Yaskawa/Motoman, Doosan) have extended their product portfolio with cobots, and new suppliers such as Universal Robots, Techman Robot, and Franka Emika have entered the market. Examples of cobots are Universal Robots' UR5 [283], Kuka's LBR iiwa [165], and Franka Emika's Panda [95] (Fig. 4.1).

The design of current cobot UIs that are on the market can be improved. Researchers have evaluated several cobot systems and found them to have low usability scores[1]. High ease of use is especially important, as the promise of cobot systems is that these will enable reprogramming by factory workers with limited programming expertise [188]. The importance of making industrial robot systems easier to use is further exemplified by several recent research projects that aim to make cobot UIs easier to use or introduce different concepts for the design of programming environments on cobot UIs [141, 156, 209, 259]. Companies are entering the market offering alternative UIs for existing cobot hardware with the proposition that these are easier to use [71], which indicates that improvement of the usability of cobots is seen as a market opportunity. Other researchers have pointed to the importance of designing cobots to be easy to use and intuitive to interact with [188, 292] and that are positively evaluated by operators, in order to make the shift from traditional manufacturing robotics to user-friendly cobot systems [183]. Therefore, we claim that establishing factors that are important for cobot design and providing cobot UI designers with a tool for usability inspection is a relevant endeavor.

In this paper, we propose design guidelines for heuristic evaluation of cobot UIs. We argue that cobot UIs require a separate, more specific set of guidelines as the context these systems operate in is distinct from other HRI (Human-Robot Interaction) or HCI (Human-Computer Interaction) applications for which heuristics have been developed previously. One of the main tasks is the programming of the cobot's task, which is a rather specific HRI application for which factors such as reuse of previous work are important. This sets it apart from applications such as teleoperation in field robotics. Moreover, cobots have a specific use, as these systems are integrated in a work context in which they are subject to interaction with operators repeatedly over long periods of time, which increases the importance of human factors and accessibility [292]. Cobots also represent a challenge in terms of the variety of users, who differ insofar as their roles, preferences, and abilities to modify the cobot control program are concerned (levels of interaction [243]). The UI should support all these users in programming, maintaining, and monitoring the cobot.

We propose a set of 24 design guidelines that can be used during the development, design, and implementation phases of cobot systems to identify UI design problems by means of heuristic evaluation. The contribution of this paper is the identification and refinement of cobot design guidelines based on literature and expert evaluation, thereby providing a resource for design and evaluation of

[1] Several usability evaluations yielded SUS scores between 50-70 for different cobot systems [85, 243], which has been argued to indicate that a product is marginal in terms of usability and should be improved [19].

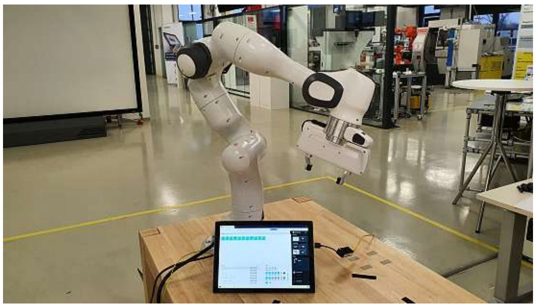


Figure 4.1: Franka Emika Panda cobot

cobot UIs (a form of intermediate-level knowledge for HRI design [180]). In Sec. 4.3, a short literature review is provided. The procedure and database search to establish the initial guidelines are described in Sec. 4.4. The guidelines were further developed by means of an evaluation study as reported in Sec. 4.5, and interviews with UI/UX design experts as described in Sec. 4.6. Finally, Sec. 4.7 contains the proposed guidelines.

4.3 Related Work

4.3.1 Interaction with cobot systems

With regard to design, several factors must be considered by a manufacturer to enter the market and be competitive. Besides the main factors functionality and safety, other factors such as risk and ergonomics assessment, risk reduction in the workspace, and specifications of use limits and transitions to other operations are required for the design of cobot applications [143]. The interaction is characterized by the levels of interaction based on safety implications of cobots. Cobots are partly completed machinery; these systems have to be designed in compliance with regulations (e.g. [270]) and include safety mechanisms [182]. Cobots have particular challenges, as they are intended to be operated in close proximity by end users with differing levels of programming experience, in the context of different types of tasks. For instance, users can take on the role of programmer, operator/supervisor, or maintainer [77, 183]. Programming and maintenance activities are usually of limited duration or even restricted to single events, while supervision requires more continuous levels of interaction and attention [183].

Another consideration is that of UI design. The cobot user interface is that part of the system that enables the user to interact with the cobot. The interaction takes place by means of system inputs and outputs, and can be realized using multiple different interaction modalities (multimodal interaction). The design of the interface determines how commands can be given to the machine or application and how information is represented to the user. Several different interaction modalities exist (e.g. visual, acoustic, haptic) and are realized in different ways (keyboard, mouse, buttons, sounds, signal lamps, projectors) (see [77, 243] for an overview). A common piece of hardware that is part of the UI is the teach pendant, which is a hand-held device that is usually equipped with physical buttons and a GUI. This is often the main interface for programming and maintenance activities [183]. Examples of GUIs are Franka Emika's web-based Desk interface [95] and the GUI of UR5, which runs on a teach pendant [283]. Teach pendants and other associated screens enable the user to program and control the cobot and to observe status information. Programming environments often contain representations of the robotic arm and its configuration, and location and trajectories of the end effector.

In order to introduce cobots in industrial environments, research has been conducted into enabling human awareness and so-called *intuitive* programming of cobots [77]. The term *intuitive* is often used to characterize UIs and refers to an effective interaction between a user and a technical system without conscious use of previous knowledge [195]. Moreover, it refers to the use of design strategies that appeal to prior knowledge in a way that makes it easier to (learn to) use an interface and reduces user effort. Another relevant term is *usability*, which has been defined as the *“extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [142]. It has been argued that making technologies more intuitive and easier to learn may promote access to jobs that typically require higher skills [188]. Developments towards making cobot programming more intuitive include integration of teach pendants and other input devices, programming by demonstration, multimodal HRI, and virtual and augmented reality (VR, AR) [160]. A drawback of using communication modes such as speech and gesture is that they add additional uncertainty; GUIs may be better suited for use in industry. Ways of programming a cobot through a GUI include making use of cobot teach pendants, Icon-based programming, CAD-based programming, and task-based programming. Another way to control the cobot is by means of haptics and force [77]. Krot and Kutia distinguish between online, offline, and hybrid programming approaches [160]. For more information on cobot programming we refer to [77, 160].

4.3.2 Evaluating cobots: metrics, measures, methodologies

Several researchers have performed usability evaluations of cobot systems. Ferraguti et al. [85] propose a methodology for executing a comparative analysis of cobots, which they apply to the cobots KUKA LWR 4+, UR5, and Franka Emika Panda. They propose that technical data as provided by the manufacturer, experimental verification, usability evaluation and required physical and mental effort to program the robot should be taken into account for such a comparative analysis. For the usability score, they used the SUS (System Usability Scale) and the Questionnaire for the Evaluation of Physical Assistive Devices (QUEAD). Schmidbauer et al. [243] evaluated three cobots UIs (Franka Emika, UR, Fanuc) using the SUS. Weintrop et al. [296] evaluated the time on task, task success, ease of use, learnability and satisfaction of the CoBlox environment they developed using Blockly, ABB’s FlexPendant, and Universal Robots’ Polyscope. For an overview of metrics, we refer to Marvel et al.’s metrology framework for Human-Robot Collaboration in manufacturing [183]. We conclude that for the evaluation of the usability of cobot systems, often qualitative and quantitative measures are combined. Examples of quantitative metrics for assessing interfaces are learning time, expert use time, and error cost, while qualitative metrics include the NASA-Task Load Index (NASA-TLX) and the SUS scale [183].

4.3.3 Design guidelines and heuristic evaluation

Heuristics are principles that can be used to identify usability problems with a particular UI. Examples of heuristics are: *"Does the interface provide feedback?"* [50] and *"Provide indicators of robot health/state"* [1]. Heuristic evaluation refers to a type of evaluation that is usually conducted by UI/UX experts, during which they use heuristics or design guidelines to identify usability problems. Quiñones and Rusu [220] describe the method of heuristic evaluation. First, each evaluator individually compiles a list of problems with the help of a set of heuristics. The evaluators combine the problems to form a list of unique problems, which are then rated for severity, criticality, and frequency by each individual evaluator. Advantages are (1) the low cost of heuristic evaluation, (2) that no extensive planning is required, (3) the broad applicability of the method, especially early in the software development process, and (4) that the method can help find many usability problems without involving end users. On the downside, heuristic evaluation requires experienced evaluators with task-specific knowledge. Additionally, the method helps identifying problems but does not offer pre-packaged solutions to these problems [220]. Heuristics and design guidelines can be considered as intermediate-level knowledge for HRI design [180].

Different sets of design guidelines and heuristics exist, many of which are targeted at the design of computer software, such as the usability heuristics proposed by Nielsen [198]. For robotics, alternative sets have been proposed for applications such as field robotics [192], robot teleoperation [1], assistive robotics [282], and HRI in general [50] (these have also been applied in video-based heuristic evaluation [298]). Robots are embodied and HRI often requires different interfaces to support multiple user roles, which sets HRI apart from HCI applications [192]. While these arguments apply to cobots as well, cobots present a different application area compared to field robotics, teleoperation, and assistive robotics. When interacting with a cobot, the human and the robot are co-located, which distinguishes this type of interaction from scenarios such as teleoperation, as operators have access to more contextual information [183] and users can, for instance, switch between physical interaction and interaction through a GUI. Cobots can be interacted with and observed via different means (e.g. programmed on a computer or using a teach pendant, or hand-guided). Cobot status can be observed on the teach pendant or by looking at the cobot itself (e.g. checking indicator LEDs). Interaction modalities such as AR, VR, and a variety of handheld devices are subjects of ongoing investigation [292]. This makes it important to guarantee consistency with respect to the way information is presented to the user across different modalities, as well as to support the user in managing their attention. Teach pendants of cobots can be rather heavy, but existing sets of heuristics (e.g. [50, 198]) do not refer to physical ergonomics, which is a relevant factor for usability of cobot systems. This means that existing design guidelines for HCI or HRI in general are not sufficient, nor are sets that have been developed for other applications such as teleoperation. In the next section, the development of a set of guidelines for heuristic evaluation of cobot systems is described.

4.4 Development of the cobot UI design guidelines

With regards to the realization of a new set of heuristics, Quiñones and Rusu [220] recommend establishing which features of the target domain are application-specific, identifying existing heuristics that can be reused, specification of heuristics according to a template, and validation of the heuristics. Common methods include collecting design recommendations from industry, from the public domain [12], and from academia [1]. The collected guidelines can then be clustered using asynchronous affinity diagramming [12] or other methods, and sorted using methods such as open or closed card sorting [1]. Methods to evaluate heuristics include performing modified heuristic evaluations, user studies based on heuristic evaluation [12, 50, 282], and expert reviews [12]. Other practices include the use of templates for problem reporting [50], customizing heuristics to the application domain [50], and using an iterative process [12, 50, 220]. We make use of these methods and recommendations to establish heuristics for cobots, as described below.

4.4.1 Procedure for establishing cobot UI guidelines

The methodology for establishing the cobot UI guidelines was as follows (see Fig. 4.2). Prior to establishing the design guidelines, an informal heuristic evaluation was conducted using existing sets of heuristics, namely [1, 198], to determine if these were sufficient or if some issues were not appropriately covered by those heuristics. In order to establish which design guidelines had previously been proposed in the academic literature, papers were collected that propose guidelines for HRI design by searching academic research databases (ACM digital library, IEEE Xplore®, and SpringerLink, Sec. 4.4.2). Design recommendations that were listed in those papers were collected and categorized during an affinity diagramming session. This resulted in clusters of guidelines that were summarized into individual heuristics. Additional heuristics and clusters were proposed on the basis of literature on cobot systems in manufacturing (beyond literature focusing on design guidelines, e.g. [77]). The preliminary heuristics were used in an evaluation of two cobot systems, after which they were revised to make them more clear, actionable, and applicable. During the first empirical evaluations (Sec. 4.5), study participants (early-career researchers with HRI expertise) were asked to apply the guidelines in the context of a modified heuristic evaluation of a cobot system, and rate the guidelines for clarity. Based on participant feedback, the guidelines were revised. Experts in cobot UI/UX design (such as UX designers at companies developing cobot systems) were invited to participate in interviews to collect feedback on the guidelines (Sec. 4.6), which led to a final revision of the guidelines. For the final version of the guidelines, see Tables 4.1-4.3.

4.4.2 Literature review with database search

We performed a database search with the aim of finding design guidelines, recommendations, and heuristics that have previously been proposed in the academic literature, especially for HRI and cobots. The inclusion criteria were that the publication had to be in English, that it contained specific recommendations and heuristics, and that its topic should concern design guidelines, heuristics, or interaction design for an interactive technology or robotics. The research question was: *What are specific design guidelines, recommendations, and heuristics that have been proposed in the academic literature, especially for HRI?*

The databases ACM digital library, IEEE Xplore®, and SpringerLink were searched for combinations of keywords. The timeframe was restricted to 1999-2019. Databases were searched for combinations of one of the terms “Human Robot Interaction", “cobot", “collaborative robot", “robot", **plus** one of the following terms: “heuristics", “design guidelines", “design recommendations", “usability guide-

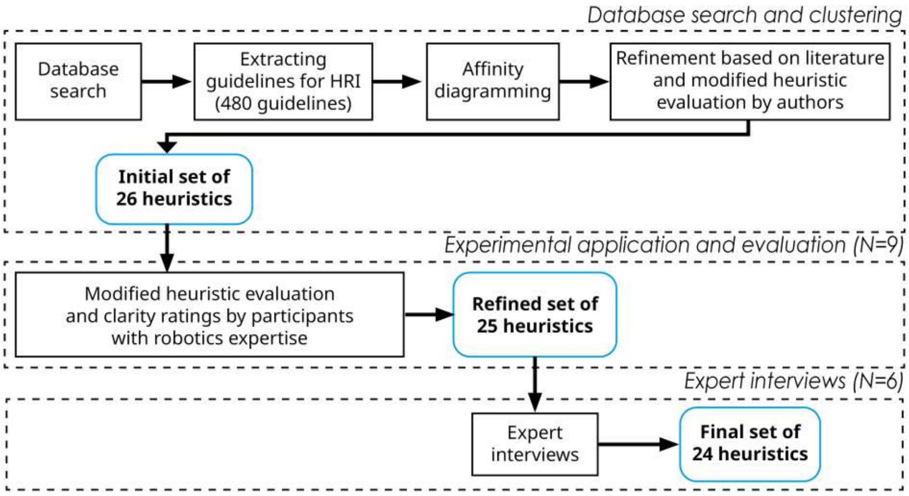

Database search and clustering
Database search
Extracting guidelines for HRI (480 guidelines)
Affinity diagramming
Refinement based on literature and modified heuristic evaluation by authors
Initial set of 26 heuristics
Experimental application and evaluation (N=9)
Modified heuristic evaluation and clarity ratings by participants with robotics expertise
Refined set of 25 heuristics
Expert interviews (N=6)
Expert interviews
Final set of 24 heuristics

Figure 4.2: Overview of the research process, in which different sources of information were used to propose design guidelines for heuristic evaluation.

lines", “design principles". If possible, it was specified that these terms needed to appear in the abstract of the paper. After searching for combinations of keywords, duplicates were removed. Then, the paper’s title, keywords, and abstract were reviewed and the paper was excluded if it did not meet the inclusion criteria. For the ACM database search (conducted 25-11-2019, 28-11-2019), 354 results were checked, resulting in 32 papers. For the IEEE Xplore® database search (on 19-11-2019), checking 2808 results led to selection of 53 papers. The Springer database search (on 3-12-2019 and 4-12-2019) yielded too many search results to check manually. For this database, we specified the search terms as a combination: ((Human AND Robot AND interaction) OR robot OR (collaborative AND robot)) AND (heuristics OR (design AND guidelines) OR (design AND recommendations) OR (design AND principles) OR (usability AND guidelines)). This yielded a total of 35,851 results. Results were sorted with SpringerLink’s “sort by relevance" feature and restricted to the first 500 items. Of these items, 51 papers met the inclusion criteria.

The database search yielded 131 unique papers after checking titles and abstracts for relevance. Each paper was checked for concrete guidelines and recommendations, which resulted in 42 papers containing 561 guidelines and recommendations. The main application domains that these papers were concerned with, were HRI in general (11 papers, e.g. [298]), rescue robotics, field robotics, safety-critical systems and tactical systems (11 papers, e.g. [1, 192]), assistive and service robotics (10 papers, e.g. [282]), telepresence systems (3 papers), and industrial robotics (2 papers, [89, 184]). After removing duplicate guidelines, this resulted in 480 guidelines, heuristics, and other recommendations.

4.4.3 Clustering of guidelines based on affinity diagramming

All 480 design recommendations were clustered into groups of similar topics by means of affinity diagramming (similar to [12]). This resulted in 38 clusters, each of which contained between 2 and 27 recommendations. Additional design issues specific to cobots were identified, mostly based on [77, 160], such as the use of multiple modalities and providing information on ways a user can teach the cobot. Finally, some clusters were merged, changed, or deleted (e.g. the cluster “cultural expectations" was deleted, as the recommendations collected in that section mostly applied to conversational agents). We summarized each cluster into a single design guideline, resulting in a total of 35 guidelines.

4.4.4 Revision based on modified heuristic evaluation

The first version of the guidelines was applied by the authors of this paper in an evaluation session that was structured as a modified heuristic evaluation modeled after [12]. The interface of Franka Emika Panda and UR5 were evaluated. The Panda cobot was controlled from the Desk environment, version 2.1.1 ced048bda3. The UR5/CB3 (version 3.9) cobot was equipped with software 3.9.1 64192 (April 1 2019), installed on a teach pendant (12 inch touchscreen) with the PolyScope GUI. A Robotiq gripper [231] plus custom 3D printed safety encasing was attached to the UR5 robot, and functionality of the gripper was embedded in the GUI with URCaps. During the evaluation, examples of applications and violations of the guidelines were collected. If no applications or violations of the guidelines were found, deleting the guideline was considered, as this was an indication that the guideline would be difficult to evaluate by means of heuristic evaluation. If there was overlap between applications and violations listed for different guidelines, we considered merging them. This resulted in 26 guidelines. Items regarding technical capabilities, human-oriented perception, and safety were removed, as these are basic requirements of robotic systems. Assessing safety or privacy is outside the scope of UI/UX design or heuristic evaluation, as safety of the cobot system must comply with (legal) standards [143].

4.5 Evaluation study based on modified heuristic evaluation

After the database search and subsequent refinement of the proposed guidelines, an evaluation study was conducted (N=9) that was based on modified heuristic evaluation of two cobot systems by participants with robotics and/or cobot systems expertise. The goal of this evaluation study was to find out if the formulations of the guidelines were clear and if they could be applied to existing cobot systems. We used feedback by participants, clarity scores given by participants to guidelines, and survey responses to revise the guidelines established in Sec. 4.4.

4.5.1 Procedure of the evaluation study

The evaluation study was structured as a modified heuristic evaluation. The evaluated cobots were described in Sec. 4.4.4. The study took place at the teaching and learning factory at TU Wien [215]. Participants were asked to sign an informed consent form and complete a personal information survey. They were explained what heuristic evaluation is, what it is used for, and they were given some examples of applications and violations of a heuristic in the context of HCI and HRI. Next, they received a safety briefing for the cobot they were going to work with. Participants were asked to read the guidelines and the survey form was explained to them (see next section). They were asked to identify applications and violations of each design guideline, to rate each guideline for clarity and to make notes if they had any other feedback. Participants were informed that their performance on the task would not be scrutinized and that constructive criticism was welcomed, as this would help to improve the guidelines. Then, they worked with the cobot by themselves and completed the survey form. Participants took about one hour to complete the evaluation study.

Participant task in the evaluation study

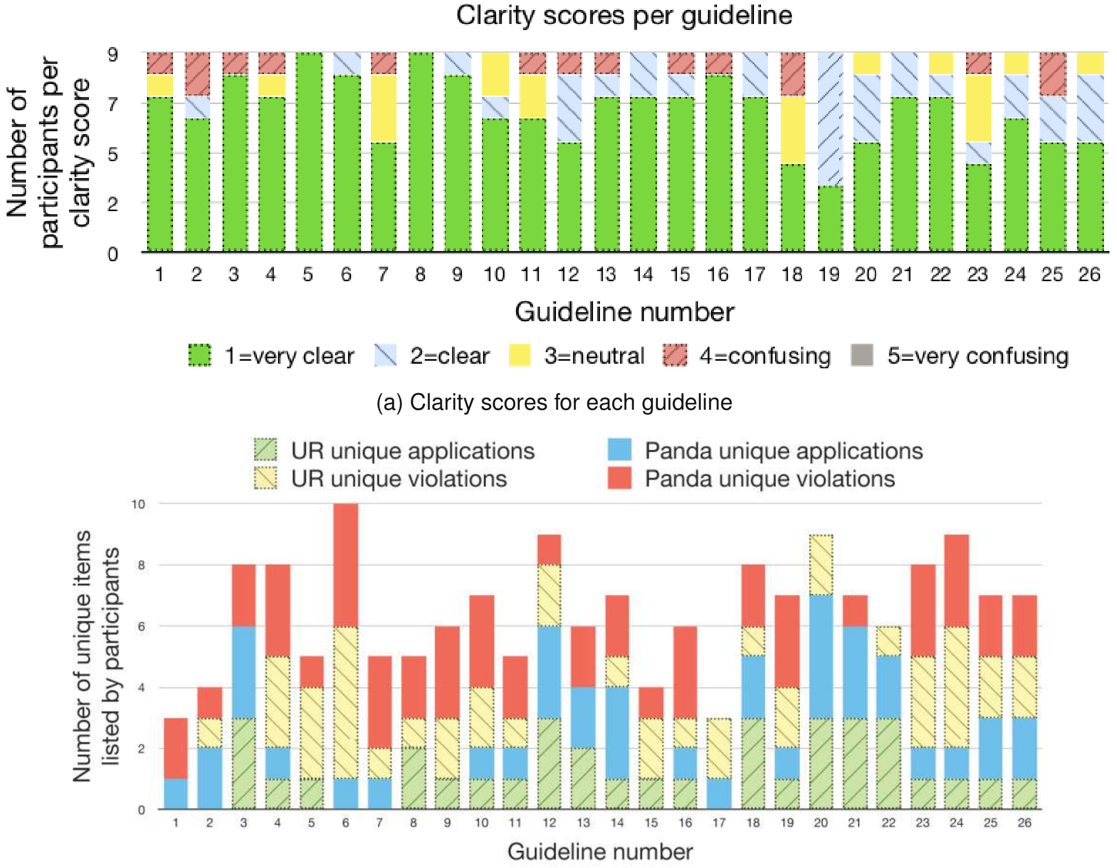
Each participant interacted with a cobot for an hour. They were supplied with a programming task, but were allowed to deviate from it. The task contained instructions to program a movement trajectory with several intermediate points for the cobot and to open and close the gripper. The task for UR5 was as follows: *1) Use the Move tab to program a trajectory. Program at least 3 points. 2) Change the second point of the trajectory you programmed. 3) Open the gripper. 4) Add another Move trajectory: Use the Freedrive button to move the robot and store some points for a trajectory. 5) Close the gripper.* The task for Panda was similar. We asked participants to complete a survey form while programming the task on which they 1) gave a rating for clarity of each guideline, 2) wrote examples of applications and violations of guidelines they encountered in the UI (for example, one participant noted *"It is not clear what is the exact location of the end effector"* as a violation regarding accessibility of information), and 3) gave feedback on guidelines. We asked for examples to check whether people could apply the guidelines.

Survey forms

The personal information sheet included questions regarding participants' professional occupation, specialization, name of employer, and asked participants to estimate their experience working with cobots in hours and robotics expertise in years. The survey form contained every guideline in the set. The first question for every guideline was *"The formulation of this guideline is very clear / clear / neutral / confusing / very confusing"*. Participants were asked to circle the answer they felt was appropriate. There were open fields where participants could list applications and violations of each guideline. They were informed that they did not need to list any examples if they did not find any, but that they could also list multiple examples. The final question on the survey form asked participants for feedback on the guidelines. Half of the participants received a survey form on which the guidelines were sorted alphabetically and the other half in reverse alphabetical order.

4.5.2 Participants

A total of 9 participants (5 male, 4 female) took part in the first evaluation sessions. Participants were recruited internally, thus 8 of 9 participants reported being employed by TU Wien. The participants were mostly early-career researchers (average age M=29.3, SD=2.2 years), with specializations in the areas of (industrial) robotics, mechanical engineering, assistive systems, HRI, and social robotics. Regarding their robotics expertise, one participant reported that their robotics experience was limited, 3 reported between 18-24 months of experience, 3 reported 4-5 years experience, and 2 reported 10 years experience. All participants had experience directly working with cobots (e.g. programming cobots).



(a) Clarity scores for each guideline

(b) This figure shows the number of unique applications and violations of each guideline were listed for both the Panda cobot and the UR cobot.

Figure 4.3: Results of the evaluation study.

4.5.3 Evaluation results

The explanation phase of the experiment, including reading the guidelines, took 20 minutes on average and the average working time was 63.3 minutes. Of the participants, 4 evaluated the UR5 UI and 5 the Franka Emika Panda UI. See Fig. 4.3a for an overview of the clarity scores for each of the guidelines. Then, the listed applications and violations were analyzed. Items were excluded if they were not specific enough, did not apply to the guideline, or were simply stated as a confirmation of the guideline. For instance, one participant listed *"Panda is very intuitive"* as a response to *Support user learning*, which was removed on the grounds of not being specific enough. Another wrote *"indicating system state"* in response to *System state awareness*, which was removed as it does not specify how the guideline is applied or violated in the UI. See Fig. 4.3b for an overview of the unique number of applications and violations for each guideline (if multiple participants listed the same application/violation, these would be counted as one unique application/violation). As multiple examples were listed for each guideline, we conclude that the guidelines could be applied to the cobot systems and that guidelines could be interpreted by the evaluation participants.

4.5.4 Revision of the guidelines

Guidelines were revised on the basis of the participants' feedback and if a guideline's clarity score was higher than 1.5 (guidelines with a score of 1 were very clear, guidelines with a score of 5 very unclear). During the revision process, the applications and violations listed by the participants on the survey form were compared to the meaning the researchers intended to capture with the guideline. Guidelines were revised by changing terms that were not understood, changing complex words to simpler formulations, removing word repetitions, and ensuring guidelines contained between 5-12 words as recommended in [12]. For example, item 12 had a clarity score of 1.67, thus *12. Errors: Accommodate troubleshooting when errors occur* was changed to *12. Errors: Give clear explanations and steps to recover when errors occur*. Terminology was made consistent (e.g. referring to the person interacting with the system as “user"). Guidelines were grouped into categories and initial category headings were added.

4.6 UI/UX Design Expert Interviews

In order to get a practitioner perspective on the guidelines and on the practice of developing cobot UIs in industry, cobot UI/UX design experts were invited to participate in interview sessions. The main aims of the interviews were to evaluate if participants could recognize and relate to the guidelines, as well as to gather feedback to the guidelines and check whether they needed revision. In other words, both the relevance of individual guidelines to participants' work practices and the formulation of the guidelines were evaluated. Six online interviews were conducted, which resulted in 41-74 min. of recorded material per interview (average 52 min.). The interviews were transcribed and coded using a thematic analysis approach [29]. On the basis of interviewees' feedback, minor changes were made to the guidelines.

4.6.1 Participants

We contacted 9 companies and company branches via email and contact forms. Additionally, 16 individuals were contacted who had cobot UI/UX design experience and/or had a leading design or innovation role at companies that developed cobots, who were found via company LinkedIn pages, referrals by university colleagues, and referrals by interview participants. We found 6 suitable participants who agreed to an interview, four of whom worked at two different companies producing cobot systems in the role of (senior or junior) UX designer or engineer. One participant worked at a university as a researcher and had prior experience in software development for cobot systems. The final participant worked at an industrial automation company and was responsible for usability testing of industrial robotic systems. Five were based in Europe and one in China.

4.6.2 Procedure and Interview Protocol

Prior to the interview, participants received the guidelines for reference, information regarding storage of data in accordance with GDPR, and an informed consent form. At the start of the interview, the interviewer (the first author of this paper) thanked the interviewee for participating, introduced the goal of the research study, and gave the participant an overview of the topics that would be discussed.

Participants were informed that constructive criticism was welcome, that they could withdraw at any time, and that they could notify the researcher if they did not want to answer a particular question. The participant was informed that the recording was started and was first asked to answer introductory questions regarding their job description, the main (design) activities and evaluation methods they used in their job, and their familiarity with cobots. Subsequently, participants were asked for feedback regarding the individual guidelines, their response to the whole set, and the categorization. Finally, participants were asked what challenges they encountered in the cobot UI design process. Participants were not informed regarding previous steps of the development of the guidelines.

4.6.3 Interview themes

A thematic analysis approach [29] was used to analyze the interviews using the software MAXQDA [185]. First, transcripts were read by both authors of this paper. One of the interview transcripts was analyzed and coded separately by both researchers (generating initial codes), who then agreed on a coding scheme, which was used to (re)code all six transcripts. The coding scheme included responses to guidelines and other categories such as design methods used. Both researchers were involved in coding the transcripts and subsequent analysis. Based on the analysis of the interviews, recurring themes were identified regarding goals, methods, and design constraints. Job-related goals mentioned by interviewees were focusing on and solving problems for the end user (explicitly mentioned by 3 different participants), improving ease of use and the user experience (2x), improving efficiency (2x), changing the minds of engineers on the team, improving system understanding, conducting research, and solving problems in the company product (1x). Methods that were mentioned most were getting feedback from team members or customers (5x), interviews with end users/customers, persona development or definition of target end users (4x), getting familiar with or explicitly evaluating competitor products (3x), comparing interfaces using metrics (such as time to complete certain tasks), identifying dependencies and requirements, drawing, and prototyping (2x).

Design constraints

Interviewees indicated that the implementation of the guidelines not only depended on the design of the cobot, but also on the application that the cobot was used for, regulations, safety measures, company culture, and differences between users. Different users need different information (mentioned 3x), have different backgrounds and levels of expertise (3x), and have different expectations. Human factors is influenced by the application and the way the work environment is organized, by regulation, and different users have different abilities (e.g. some users are color blind). Cultural effects and country-specific regulations will also influence how the guidelines can be implemented. One person mentioned that autonomy given to workers by the companies varies greatly from one country to another. While we would propose that information that is necessary to (better) achieve the user's task should not be hidden by the system manufacturer, we note that company culture, the culture of the region, and commercial interests will influence how guideline number 3 (in Table 4.1), accessibility of information, is implemented in practice.

Diversity of end users as an effect on guideline implementation

Interviewees emphasized that there are many different end users for whom they are designing interfaces. People who interact with cobots have different roles in the company (such as operator or programmer). The type of interface that is desirable depends on the previous knowledge and expertise of the end user, but also on what kind of application the user wants to implement on the cobot. A software developer with several years of coding experience, but no previous interaction experience with any robot needs different user guidance than a shop floor worker who has several years of experience in working with machines and industrial robots but has never programmed or controlled them by themselves. Novices were often contrasted to expert users during the interviews. In addition, different stakeholders of the cobot development process were mentioned, such as company salespeople, system integrators, customers, end users, engineers, and programmers. We connect the fact that different users have different needs to the frequent use of methods such as developing personas. While many of the interviewees emphasized that their focus is on the end user, they also noted that it can be difficult to get access to end users. One participant mentioned that in specific industries, exchange of information is not desired by companies, although this information is much needed to improve cobot UI/UX design.

Response to design guidelines and revisions

Some participants responded to almost every guideline, either by stating a confirmation, supplying a practical example, asking for clarification, or voicing disagreement. Guidelines *13. Human factors* and *17. Adaptable system architecture* were responded to by most participants (5 out of 6). Participants also remarked on the whole set, its potential uses, and on the categorization of the guidelines. One participant noted that *"it is something I can relate to, even though I never had this structured list before (...) it's (...) putting a word on most items I do here at work"*. Two participants explicitly stated they thought the guidelines were quite complete. A few extra topics were suggested, such as safety (which we had previously decided to exclude) or learning about safety (which is implicitly included in item 9). Possible alternative uses of the guidelines were mentioned, such as using the guidelines as a checklist, using them for discussion with top management to indicate which items need to be worked on, or applying the guidelines to other machine interfaces in a manufacturing context. Two participants remarked that similar UI/UX guidelines exist, which is understandable as the guidelines are based on literature search. Participants made comments that many of the issues touched on by the guidelines were large research projects in themselves. One participant also emphasized that cobots are part of larger systems, and that programming cobots is just one piece of the puzzle when it comes to solving automation problems.

Interviewee feedback indicated that the guidelines did not need substantial revision. Details in the wording were adjusted, for instance *"operator attention"* was adapted to *"user attention"* as this guideline did not apply only to operators on the shop floor. Several guidelines were adapted to reflect remarks by participants. For instance, the item *accessibility of information* was adapted, as participants pointed out that some information is not understandable to end users (such as raw data), and information requirements and accessibility also depend on the user, culture of the region and company, and user access. Guidelines 3, 8, 9, 11, 13, 15, 17, and 18 were adapted slightly, 20 was adapted substantially for clarification and merged with a different guideline, and several initial category headings were revised based on participant feedback (the numbering refers to the guidelines as presented in Table 4.1-4.3).

4.7 Proposed set of guidelines

The proposed set of cobot UI design guidelines consists of 24 items, see Tables 4.1-4.3. An explanation and an example are included for each of the guidelines. The guidelines cover the main topics of situation awareness, system understanding, task efficiency, human factors, configurability, and interaction design of the UI, with a focus on the usability and user experience of the cobot system. The guidelines should be considered from the perspective that the user group is diverse and that different users will have different needs, for instance in terms of the information they need to achieve their task, that they might need different levels of support and complexity, and have different abilities, all of which have consequences for the required interface design. Regarding intuitive use (guideline 22), mental models (guideline 7), and user learning (guideline 9), we remark that the differences in users' backgrounds will also have an influence on how these guidelines are operationalized; for different people, different things are familiar, and (thus) different UI features will be easy to use or easy to learn.

4.7.1 Comparison with existing heuristics and usability guidelines

After the final revision based on the expert interviews, we compared our proposed guidelines with heuristics for robot teleoperation [1], assistive robotics [282], and HRI [50]. Guidelines 2, 5, 7, 9, 11, 12, 15-18, and 20, as proposed in the current paper are not included in the set for assistive robotics [282]. While the compact form of the HRI heuristics [50] is useful for heuristic evaluation, not all issues indicated by the guidelines proposed in the present paper are covered (guidelines 5, 9-13, 16, 18, 24, and to some extent 23 are not covered by [50]). The set that was the most similar to the guidelines proposed in the current paper was the one for teleoperation [1]. We note that while there is some overlap of many guidelines on a surface level regarding the topics, the formulation and meaning of the guidelines is different. Several guidelines we propose for cobots are not included in the teleoperation guidelines, namely guidelines 12 (reuse of previous work), 13 (human factors/ergonomics), and 19 (consistent behavior), due to the differences between the domains of robot teleoperation and cobots for manufacturing. Several guidelines had overlaps, but with different practical implications. For example, guideline 20 (Multimodal UI) is in a sense similar to the item *"Complement video stream with feedback information from other sensors"* [1, pp. 257-258], but the presented information is very different (video stream from remote teleoperated robot versus a programming environment on a GUI in a manufacturing context). Some items relevant for teleoperation are not relevant in the context of cobot systems (e.g. *"Ability to self-inspect the robot's body for damages or entangled obstacles"*[1]).

4.7.2 Validity and generalizability of guidelines

In this section, potential threats to validity are considered. In the evaluation study (Sec. 4.5), participants worked in a factory-like setting with equipment and robots [215]. This gives the experiment some level of ecological validity (findings generalize to the real world [131]). The first evaluation study indicates that guidelines can be applied to cobot systems. The external validity of the study is

Situation awareness
1. System state awareness: Inform the user on the cobot's state. The interface should support the user in maintaining appropriate awareness of the system's state. *Example: The cobot informs the user via lights on the robotic arm about the current state, for instance a green light means a program is running.*
2. Situation awareness: Inform the user regarding the cobot's environment and configuration. Help the user in understanding the configuration of the cobot in its environment, as well as other sensor inputs. *Example: The GUI shows a 3D model of the robot, which indicates the cobot's configuration and end effector position.*
3. Accessibility of information: Allow users to access information required for the task. Make sure the information the user needs for the task is available and accessible, considering possible restrictions. *Example: When editing a trajectories, a user can access information about the exact end effector location on the GUI.*
System understanding
4. Feedback: The UI is responsive to user actions. Respond to user actions so the user can follow task progress and understand the effects of their actions. *Example: The UI is responsive when buttons are clicked, when the user navigates to a different menu, or when values are updated.*
5. Affordances: Signify how the user can interact with the cobot. The interface should indicate which actions are currently possible and which ones are not. *Example: An icon of a trash bin next to a stored point indicates the point can be deleted by clicking on the icon.*
6. Errors: Give clear explanations and steps to recover when errors occur. Tolerate minor user errors, prevent critical system errors, support undo and redo. *Example: When a trajectory cannot be executed, a popup appears with an explanation why the error occurred and steps to recover from the error.*
7. Mental model: Support the user in understanding the way the system works. Support the user in understanding the connection between user actions and system response, for instance by providing feedback and using appropriate terminology. *Example: The user can play a programmed trajectory as an animation on the GUI before execution by the cobot.*
8. Help and documentation: Provide contextual help and documentation. Give users clear explanations of functionality and errors. *Example: When the help icon is clicked, the help menu displays help items that relate to the functions that are currently on the display.*
9. Support user learning: Help the user solve their (automation) problem. Support trial-and-error behavior and provide templates, contextual instructions or other clues that indicate how the cobot can be interacted with. *Example: Templates for robot tasks are provided, so the user has an idea what a program should look like.*

Table 4.1: Guidelines for heuristic evaluation of cobot UIs, continued in Table 4.2.

Task efficiency
10. Efficiency: Avoid unnecessary work on the user's side. Minimize the number of steps required to achieve goals and provide shortcuts. *Example: The user does not have to set the speed and acceleration for each point in a trajectory, but can specify these values for the whole trajectory.*
11. Task progress: Communicate to the user which task is being executed. The GUI should make it easy for the user to follow task execution by indicating previous, current and next steps. *Example: When the robot is executing a series of actions, the current action that is being executed is highlighted on the GUI.*
12. Reuse: Enable reuse of previous work. Support users in reusing their work or the work of others. *Example: Previous programs can be copied and edited.*
Human factors
13. Human factors: Design cobot and UI with ergonomics and accessibility in mind. Ensure the cobot UI is comfortable to work with for the necessary duration. *Example: The teach pendant is light to carry or can be placed on a table, so it will be comfortable to work with it for a few hours.*
14. Avoid cognitive overload: Reduce mental strain. Support recognition instead of requiring users to recall information, and limit the number of options that are presented. *Example: A function name such as "Wait" is easy to remember and indicates its function.*
15. User attention: Support the user in directing their attention. Make menu items that need attention visually salient. Do not attract attention unnecessarily. *Example: When the user is in the submenu for editing a trajectory, the menu items for editing points are the largest items.*
Configurability
16. Level of automation: Let the user determine the level of human input. The user can decide to integrate human input or to make the program fully automatic. *Example: There is a function for integrating human input, which requires the operator to press a button before the robot continues its task.*
17. Adaptable system architecture: Enable easy software integration after hardware exchange. The system architecture should allow for adapting the system to different types of tasks and application scenarios. *Example: It is easy to exchange the gripper and add sensors to the system.*
18. Adaptable tasks: Support easy editing of robot programs. Robot programs, trajectories, configurations should be editable by the user. *Example: Points in a previously stored trajectory can be deleted or changed.*

Table 4.2: Guidelines for heuristic evaluation of cobot user interfaces, continued in Table 4.3.

Interaction design of the UI
19. Consistent behavior: Make sure cobot and UI behave in a consistent way. Cobot behaviors, movement, and responses are predictable. *Example: The cobot always executes the same motion trajectory the same way.*
20. Multimodal UI: Consider the relation between different interaction modalities. Manage user attention across modalities and ensure the way information is presented via different modalities is consistent. *Example: The system provides feedback with LED lights on the cobot, which matches specific events on the UI.*
21. Graphic design: Design GUI items with usability, accessibility, and aesthetics in mind. Make sure information is presented in a clear and structured way, and use color, contrast and salience appropriately. *Example: Fonts are legible and the interface has appropriate contrast.*
22. Clarity of interface: Ensure the UI is easy and intuitive to use. Avoid a complex UI design; make use of simple graphics and icons. *Example: When selecting an action for the cobot, a sub menu for editing this action opens automatically.*
23. High vs. low complexity: Display programming functions at different levels of detail. Allow users to switch between simple and more complex ways of programming the cobot. *Example: There is the possibility to change between a simple version of the UI and a more complex version that provides more options.*
24. Customizability: Support user preferences. Enable users to change the interface according to their wishes and needs. *Example: It is possible to adapt different features based on user preference, such as the size of windows on the UI.*

Table 4.3: Guidelines for heuristic evaluation of cobot user interfaces.

threatened to some level as the participants did not have specific expertise in applying design guidelines. Ideally, participants would have known how to work safely with cobots, would have previously worked with heuristics, and would have design-related expertise, besides having time to participate. As the most important aspect was being able to safely program the cobot, we chose robotics expertise as most relevant criterion. We chose to conduct the interviews with participants who had expertise in cobot UI/UX design to balance this initial focus with people with design-related expertise. Participants in the expert interviews had experience with other cobots besides those used in our experimental evaluation. This provides indications that the findings generalize to other cobots (external validity). Feedback by participants given in the studies may be biased by their desire to cooperate, which may have led to feedback that was skewed on the positive side. We aimed to lessen this effect by specifically asking for (constructive) criticism and by using the clarity scores in the evaluation study as an indication of those items that were most unclear and that thus needed revision. The sample size was small for the both the evaluation study and the interviews. In both studies, the main aim was to obtain detailed feedback to be able to improve the proposed guidelines, which was achieved. Next, quantitative evaluation can be used to validate the guidelines. Such an evaluation study could also evaluate the use of heuristics that comprise a shortened, summarized version of the presented guidelines, which could lead to possible additional refinement.

4.8 Conclusion

In this paper, a set of 24 design guidelines for the heuristic evaluation of cobot UIs is proposed. Three different sources of information were used, namely a database search, modified heuristic evaluations with participants with robotics expertise (N=9), and interviews with cobot UI/UX design experts (N=6). On the basis of a comparison with existing design guidelines for HRI, we note that the proposed guidelines are specific to cobot UI design in a manufacturing context and that they are distinct from existing guidelines in the literature.

Visions of future manufacturing in which human workers can reprogram cobots in collaborative interaction settings [243] can only be realised if cobot technologies are sufficiently usable. The design guidelines that were proposed in this paper indicate which design features are important in such interaction scenarios. However, more research into cobot design and design proposals to improve current cobot UIs are necessary.

Acknowledgements

Supported by Doctoral College TrustRobots, TU Wien. The first author of this publication was responsible for the research project and writing of the publication, and was supported by the second author. We thank Hans Küffner-McCauley, Tanja Zigaert, Michael Hornáček, all paper reviewers, evaluation study participants, and interviewees.

CHAPTER 5

Programming Robot Animation Through Human Body Movement

This dissertation chapter is reprinted, with permission, from: Frijns, H.A., Stoeva, D., Gelautz, M., Schürer, O. (2024) Programming Robot Animation Through Human Body Movement. In: Proceedings of the 2024 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO) © 2024 IEEE. See also the copyright statement in A.5.1.

5.1 Abstract

The aim of our work is to make it easier for non-programmers to program robot motion. We designed a system that translates human motion to motion of a (virtual and physical) humanoid Pepper robot. The system enables programming robot animations using pose-matching imitation of human motion. We developed a prototype that implements recording functionality, evaluated the prototype, and revised the system architecture based on feedback from participants with expertise in dance or programming the Pepper robot. We extend a conceptual design space for computer animation by including a physical robot. On the basis of this design space and the prototype development, we describe interaction design considerations for robot animation systems that implement human-humanoid imitation systems.

5.2 Introduction

Human-Robot Interaction (HRI) is a multidisciplinary research field, necessitated by the complexity of developing interactions with robotic systems taking place in human social space. It has been argued that HRI research has thus far been dominated by laboratory studies with simplified views on social interaction [150, 56, 312]. Inclusion of domain experts on social and nonverbal behavior such as interaction designers, social scientists [56, 216], and expert dancers [147] has been proposed as promising for developing robotics applications for social settings [56, 216]. Researchers have advocated for inclusion of animators and dancers in robot motion design, due to their expertise on making characters seem lifelike through motion [18, 155], audience perception, and human movement features [168]. Sirkin and Ju [251] argue that embodied design improvisation can help designers of physical interactions with everyday objects (including robots) bring out tacit knowledge, that is, knowledge arising from the body and its situated interactions. However, experts with knowledge required to develop social interactions may lack robotics and programming expertise [56, 216].

To enable robot programming by domain experts such as dancers, systems and input methods with a high ease of use are beneficial. Our focus is on systems that produce robot motion for social HRI and dance, using imitation of human motion as a programming tool. We developed a prototype for animating robot motion that implements a human-humanoid pose-matching imitation system, as described in Sec. 5.4. The prototype uses Kinect v2 to realize imitation of human motion by the Pepper robot, with a graphical user interface (GUI) to start and stop the motion recording process. An evaluation of the prototype led to a revision of the system architecture. In Sec. 5.5, we outline interaction design considerations for robot animation systems, especially in relation to the development of our prototype. We extend a design space for computer animation [293] for robot animation. The contributions of this paper are this extended design space and interaction design considerations for human-humanoid pose-matching imitation systems, along with an implementation and evaluation of a prototype.

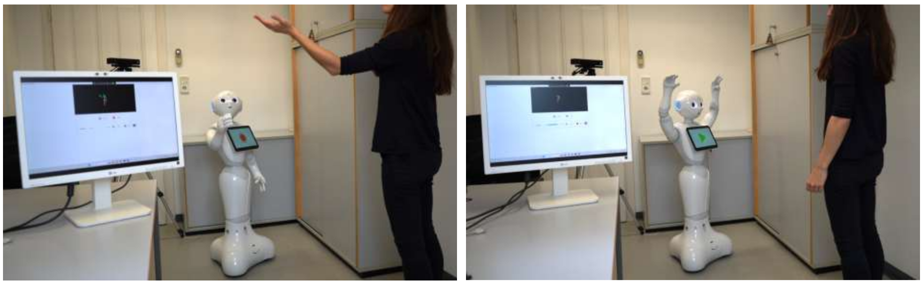
5.3 Related work

Reasons for animation of robot motion include supporting human-robot communication, generating convincing behavior, suggesting emotion, animacy [245], or personality [227], expressing information such as internal state, intentions, and attitudes, or coordinating joint activity [129]. Our focus is on expressive gestural motion. Expressive motion serves purposes such as communication, and its meaning is influenced by contextual factors [288]. Gestural motion involves gestures and behaviors that indicate a robot's state and convey information, and can be performed by executing a series of poses [110]. Techniques for robot motion design include 3D animation studies, developing a skeleton prototype of the robot, Wizard-of-Oz (WoZ) exploration of movement possibilities, and video prototyping [129]. Tools for realizing expressive robot motion need to guarantee consistent playback, safety, and scalability, and balance the complexity of information presentation while enabling access to useful information [110]. Saerbeck and van Breemen [239] distinguish three classes of robot motion design methods: trajectory design methods, motion editing methods, and high-level behavior design methods. Examples of input methods for robot motion with high ease of use can be found in the domains of Programming by Demonstration (PbD) (also referred to as Learning from Demonstration (LfD) or imitation learning [222]) and robot teleoperation (e.g., [2, 235]). PbD is a technique in which a human shows a robot example behaviors [33] that are used to generate a robot program. Bravo et al. [33] describe PbD strategies: manipulation of a robot body (kinaesthetic teaching or teleoperation), manipulation of a physical object or a (virtual or physical) robot representation, demonstration using a GUI or user body movements, and multimodal demonstrations. Similarly, input methods for LfD include kinaesthetic demonstrations, teleoperation, and passive observation [222]. These methods differ in terms of their ease of use, possibility of application to systems with a high number of DOF (degrees of freedom), and ease of mapping demonstrations to the robot.

Motion imitation systems using Kinect have been developed for the NAO [8, 311] and Pepper [96, 151, 126]. Our work differs from this, as our focus is not on the imitation, but rather on the interaction design of a system for robot animation that implements imitation. Several authors describe choreographing or designing robot motion with a GUI, e.g., for the HRP-4C robot [194], enabling dancers to give feedback to CoBot robot motion [155], or observing how animated motions combine with procedural layers [227]. Our work differs from GUIs for robot animation, as we focus on the interaction design of systems that are distributed in space, involving, besides a GUI, a humanoid robot, sensor(s) for detecting human joint positions, and an end user controlling the robot with body movement. Depth sensors such as Kinect have been used for computer animation of virtual characters [173, 208, 241]. Our work focuses on motion for a physical, co-located robot instead, which has consequences for the spatial setup of the system and how user attention is distributed across this space. Several works implement puppeteering a robot using a GUI, e.g., for the desk-light shaped AUR robot [130], the customizable Blossom robot [263], or a WoZ interface for controlling Pepper [228]. Balit et al. [18] propose software for editing robot motion and use the robot Poppy Torso to record poses. Wölfel et al. [302] developed the ToolBot system for making reproductions of handicraft with the KUKA LWR IV robot arm and a GUI for editing motions recorded with motion capture. Our work differs, as in our prototype the user's whole body is a form of input for animating a humanoid robot. Work that is most similar is that by Porfirio et al. [216] and work implementing Extended Reality (XR, an umbrella term for methods such as VR, augmented reality, and mixed reality) for HRI software tools as outlined by [9, 55]. Porfirio et al. [216] developed the system *Synthé* to program social interactions for the NAO robot. This system enables designers to use bodystorming, a technique in which they use their bodies and props to brainstorm about the interaction. We similarly focus on prototyping using the human body, but their focus is on programming interaction scenarios and only pre-programmed animations are used. Coronado et al. [55] review research on HRI software tools that implement XR devices, game engines, physical robots, and sensors for human input. For example, Alonso et al. [9] outline a system for VR teleoperation of the NAO robot. Here the user views a virtual robot using a VR headset, which differs from our setting in which the physical robot can be observed by the human interaction partner at all times.

5.4 Recording motion for animation

Our focus is on systems for robot animation using a human-humanoid pose-matching imitation system. A prototype was developed that enables programming the robot Pepper by means of human motion demonstrations with Kinect v2 input. The system's imitation functionality translates a human's body pose to joint angles for Pepper, so that it directly and continuously imitates the human's body pose. The system can be used to develop robot animations: the user starts the recording, demonstrates a motion that is directly imitated by the robot (Fig. 5.1a), stops the recording, and can then replay the recording for execution by the robot. During replay (Fig. 5.1b), the robot is not imitating the current body pose of the user, but executing recorded poses. The system needs to implement functionalities to store and replay recordings. Moreover, the user has to be able to infer system status and give commands to change it. Recording human motion for replay on the robot can be seen



(a) User interaction with the system during recording (after revision, see Sec. 5.4.3). The screen (left) displays the recording GUI. Pepper's tablet displays an icon indicating current system state.

(b) Interaction during replay, when the robot executes recorded motion.

Figure 5.1: User interaction with the system.

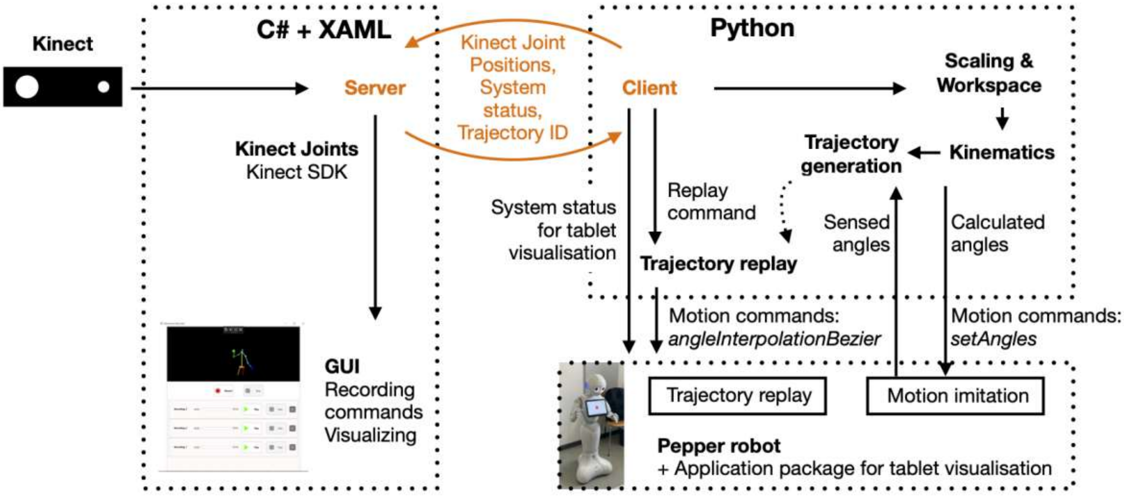


Figure 5.2: Overview of the revised system architecture. Kinect input data is processed and sent to a Python program via a websocket. On the Python side, human joint positions are mapped to Pepper's workspace and a kinematics module is used to calculate Pepper's joint angles. During direct motion imitation, these joint angles are sent to the Pepper robot as motion commands. When recording, joint angles are sensed and stored for later execution.

as a form of robot programming. Target groups include users with little to no programming skills, for instance in the fields of education and entertainment, or HRI researchers and others who want to quickly prototype robot motions for use in interaction scenarios. An iterative design process was followed in which the prototype was evaluated (Sec. 5.4.2) and then revised, resulting in the system architecture in Fig. 5.2.

5.4.1 Technical implementation

The prototype was developed for the humanoid robot Pepper by Aldebaran, which has 20 DOF [254], and Kinect v2, with code for human-humanoid motion imitation running on a laptop. A GUI provided functionality such as making recordings and replaying them. The input to the program consists of 3D human joint positions; output consists of robot motion commands. During imitation, joint angles for Pepper are calculated in near real time, and during replay the robot executes motion based on recorded human joint position data.

Performing visual demonstrations of user body movements suffers the drawback of the correspondence problem, as there is a substantial difference between human and robot bodies [33, 222]. Human bodies have different motion possibilities and constraints compared to robots, which means that devising a mapping is required; Pepper's arms have fewer DOF than human arms and the robot does not have separate legs but instead one 'leg' attached to a wheeled base. We chose a mapping to achieve pose similarity rather than a mapping between end-effector positions.

The imitation was realized by capturing human joint positions using Kinect v2 skeletal tracking with a C# Visual Studio project. The imitation system allows for controlling Pepper's ShoulderPitch, ShoulderRoll, ElbowYaw and ElbowRoll joints of both arms, the HipRoll, HeadPitch and HeadYaw joints, and the opening and closing of both hands. In our system only the person closest to the sensor is tracked and only a subset of Kinect frames is processed. Using C# with an XAML project and the Kinect SDK, human joint positions were visualized on the recording GUI. A websocket was implemented that sent human joint positions to a Python program. Human joint positions were scaled to fit Pepper's limb lengths and mapped to its workspace, to ensure a reachable position while maintaining the pose configuration. We developed an inverse and forward kinematics module [260] to calculate the required joint angles for Pepper to achieve a human's pose. The NAOqi Python SDK [254] was used to generate motion commands for Pepper.

5.4.2 Intermediate evaluation for system development

A user evaluation was conducted with six participants who either had dance/movement expertise or experience programming Pepper. Pepper programmers are familiar with the robot and can compare the use of our system to their experiences programming Pepper. Movement experts are one of the target groups, and the system needs to be easy to use for them. Movement expert participants (2 women, 1 man, age M=35.3 years, SD=3.0 years) had occupations such as movement or dance teacher, choreographer, and dancer. Self-reported programming experience scores (on a scale from *1=very inexperienced* to *10=very experienced*, question based on [82]) were 2, 3 and 4. Self-reported familiarity with robots scores (on a scale from *1=not at all familiar* to *5=very familiar*, question based on [27]) were 1, 2 and 3. Three participants, students in the fields of informatics and electrical engineering, were invited who had experience programming Pepper (2 women, 1 man, age M=27.3 years, SD=3.2 years). Self-reported programming experience scores were 8, 9 and 9. All three rated their familiarity with robots to be 5 out of 5 and reported having programmed Pepper using Choregraphe, Python, C++ and/or ROS wrappers.

Participants were given informed consent and data consent forms to sign. Then, participants interacted with the imitation system for 5-10 minutes to explore its functionality. Participants were asked

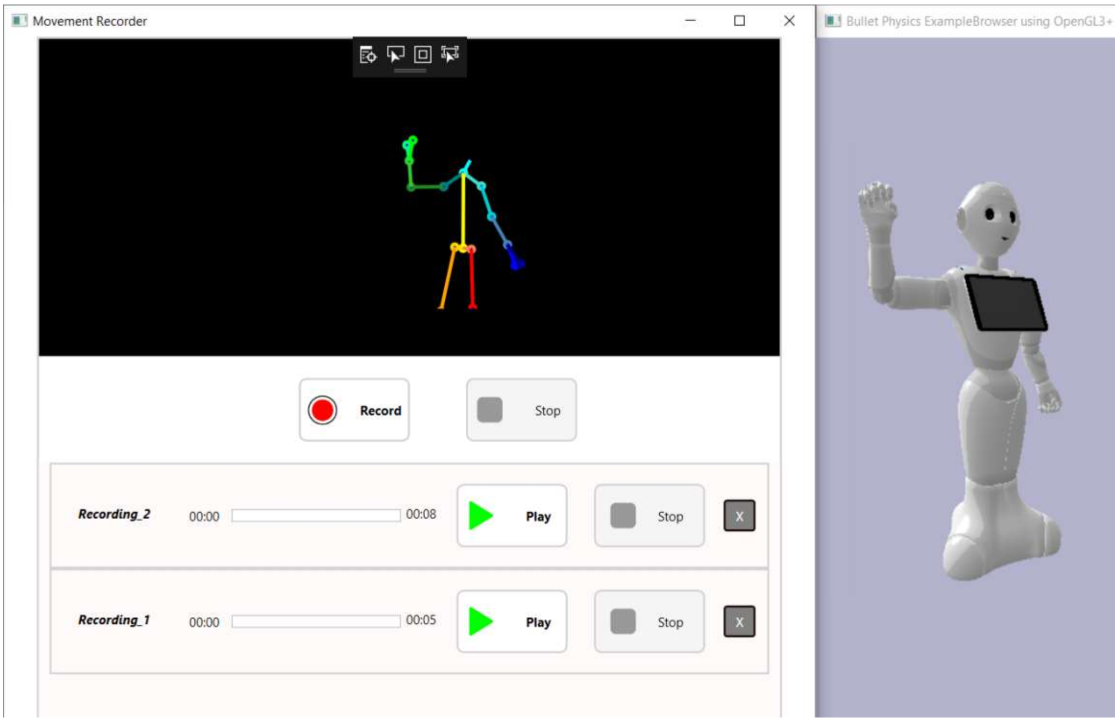


Figure 5.3: GUI and qiBullet simulator executing the same pose.

to record three motions that were developed by a dance professional and that covered different kinematic chains and gaze directions. The robot would imitate the participant's motions continuously, except when replaying a recording. Participants were interviewed on their experience of the system. All participants completed the task of recording and replaying the gestures using the GUI. The System Usability Scale (SUS) rating of the system with recording GUI by the movement experts was M=72.5 (SD=10.9), the programmers' rating was M=87.5 (SD=11.5). According to Bangor et al. [19], systems with a SUS rating above 70 are acceptable in terms of usability while better products score in the high 70 to 80 range.

Interviews were transcribed and analyzed using a thematic analysis approach [29] using MAXQDA [185]. The convention we use in the following is that P1 indicates *participant 1 with experience programming Pepper robots* and D1 indicates *participant 1 with dance/movement expertise*. Across both groups, participants commented on the recorded motion, the leaning motion that was part of the recording, the motion imitation, motion speed and other motion qualities, and responded to the whole system and participant attention during the task. Regarding imitation accuracy, responses included that fast motion looks fine but slow motion may need to be smoothed as it looked robotic or jittery (P1, P3), and that imitation worked very well (P2, P3), experiencing imitation as intuitive (P2). D2 had the impression that it seemed easier for the system to replicate fast movements than slow movements. There were some differences between the responses of both groups. Programming participants responded to the Kinect and perceived sensor inaccuracies, responded to individual features on the GUI, and suggested potential improvements for the GUI. Movement experts responded to the robot,

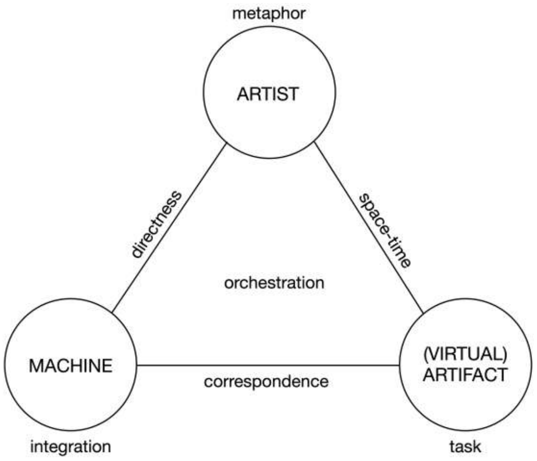


Figure 5.4: Design space for computer animation from Walther-Franks and Malaka [293].

compared its motion to human motion possibilities, remarked on a sense of togetherness, compared the relation between human and robot in this setup to a pedagogue-pupil relationship, and indicated feeling a sense of empathy.

5.4.3 Revised System Architecture

After the evaluation, the robot's motion in response to the user, the method of trajectory generation, and the GUI were changed. Fig. 5.2 shows an overview of the revised system architecture. Recorded motion can be translated into a trajectory by either using calculated robot angles that are sent as motion commands or robot angles that are sensed during motion imitation (angles that are actually reached), and choices need to be made regarding the sampling rate. To prevent recalculating angles for every replay, and for personal data protection considerations, the system architecture was adapted to store robot trajectory data. Specifically, the trajectory is now based on stored sensed robot angles, so the trajectory during replay of motion is more predictable; the sensed angles have been seen by the user during recording. To realize smoother trajectory replay, a trajectory was generated on the basis of sensed robot joint angle data using the `angleInterpolationBezier` function from the NAOqi ALMotion module for robot joint control. This is a blocking call, as opposed to the non-blocking `setAngles` command that was used previously.

The GUI was adapted to have boxes for separate recordings, and a virtual robot was added using qiBullet [36], which implements a physics simulation of the Pepper robot (Fig. 5.3) to enable programming without direct execution by a physical robot. We added visual feedback on the tablet regarding the robot's status (e.g., a recording icon).

5.5 Interaction design considerations

In this section, we outline interaction design considerations, using Walther-Franks and Malaka's design space for computer animation [293] (Fig. 5.4). We use this design space and extend it, to em-

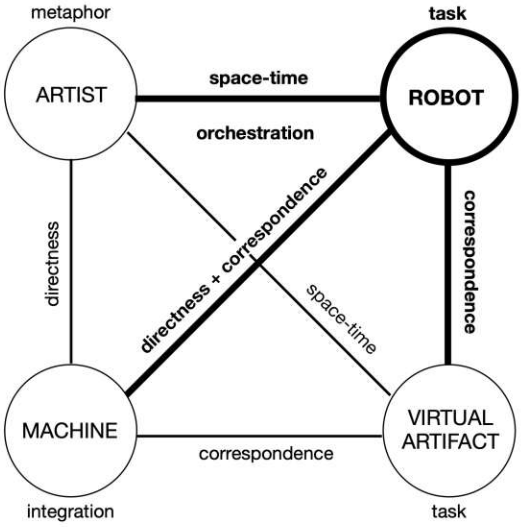


Figure 5.5: Extended design space for robot animation based on [293] with the additional item Robot.

phasize how the introduction of a physical robot platform complicates a computer animation task, requiring consideration of additional interaction design aspects. We extend the design space (Fig. 5.5) to describe systems for animating humanoid robot motion using imitation of human body movement. The physical presence of the robot and other devices that are distributed in space affect the way the user will divide their attention across these devices.

5.5.1 Design space for computer animation

In the design space for computer animation by Walther-Franks and Malaka [293] (Fig. 5.4), the entities involved in interaction include the *Artist* (who designs the motion), the *Machine* (combination of hardware and software), and the *Artifact* (the moving virtual character). The *Task* refers to what the tool is used for, namely animation. *Integration* refers to the way the input device is used as a means of control of character DOFs, and involves making decisions regarding which joints are mapped. *Correspondence* refers to the match between input entering the input device and the resulting animation of the artefact. *Metaphors* mentioned in [293] include the conversation metaphor, manipulation metaphor, and embodiment metaphor. *Directness* refers to the spatio-temporal distance or offset between user and animation, and includes issues such as precision and occlusion. *Orchestration* refers to how information is presented to the user and which actions the user can take in a particular order. *Space-time mapping* refers to the mapping of the dimensions of time and space from input to output. For instance, video playback involves a mapping from time to time, while computer puppetry involves a mapping from space-time to space-time [293].

5.5.2 Extended design space for robot animation

In this section, we highlight interaction design considerations for robot animation systems that implement a human-humanoid imitation system, based on our prototype, using the extended design space in Fig. 5.5. In systems for robot animation, the *Artifact* is split into a *Virtual artifact* (the recording data and the animation of a virtual robot) and animation of the physical *Robot*. Introducing a *Robot* adds additional *Correspondence* relations and affects *Orchestration*.

Task

Main animation tasks include motion creation, motion editing, and motion viewing [293]. For motion creation, choices are to be made regarding the type of demonstration (trajectory demonstration or keyframe demonstration), which data to use to generate a trajectory during replay, the sampling rate, and speed of the robot's motion. These choices have several implications for the experience of the interaction, and should be made in relation to the aim of the system. Choices may affect, for example, the required interaction during motion demonstration. A different user interaction is required when the user can demonstrate a motion fluently, as compared to an interaction in which the user demonstrates a series of poses for keyframe demonstration (which the system will interpolate between). Choices may also affect if timing information for the motion is recorded by default or not, how predictable motion will be during replay, and how detailed recorded motion will be. The sampling rate has consequences for the smoothness and precision of the motion. The sampling rate can be made dependent on the motion speed of the robot, meaning that for fast motions, more frames per second are stored to preserve detail. Recorded motion can be observed on a physical robot and/or a virtual robot. There are different ways to view the motion; viewing the motion while it is connected to the human demonstrator, versus viewing the motion when replaying a recording, disconnected from the human demonstrator. One consequence is that the robot should safely be moved from its current configuration into the first recorded pose when executing a recorded motion. For motion viewing, the system needs to take the starting pose into account, and move the robot from its current configuration into the first recorded pose. This is especially relevant for working with a physical robot platform, as the robot may otherwise jolt into the first pose. Potential extensions of the prototype for motion editing include trajectory editing capabilities (for individual joints). For generating motion that is both expressive and goal-oriented, it may be necessary to combine different types of motion, parametrize motions, or integrate additional objectives when the trajectory is generated (see [110, 153, 238, 227]).

Part of the *Task* involves interaction with the GUI, which can include, e.g., visual representations of system status, sensor information, existing recordings, feedback to user action and feedback regarding mode changes. Providing a stick figure that visualized human joint positions detected with Kinect gave participants information regarding tracking errors and accuracy (P1, P2). A virtual robot can be included for replay of the motion, troubleshooting in case errors arise, and working without the physical robot present. Providing options to associate additional data to recorded motions can be useful for organizing and reusing recorded motion. For instance, recordings could be named, labeled with various metadata, and visual representations of recorded trajectories may be stored alongside. As suggested by P3, motion commands could be exported to a Python script, to give end users the option to use the code file with other software

Directness

Directness depends on the sensors, calculation speed, network delays, and speed of robot motion execution. Executing motion commands by a wireless connection to the Pepper robot may add latency as compared to a virtual robot [126]. Use of (a single) Kinect can lead to occlusions and inaccuracies in the tracking data. Other possible input methods include, e.g., different depth sensors, using a combination of RGB(D) cameras to reconstruct the 3D scene, or use of on-body sensors. Each method has its (dis)advantages. Combining several sensors from different viewpoints may make tracking more accurate, but may complicate the work process and introduce costs for additional sensors.

Correspondence

We can consider the aspect of *Correspondence* on several levels, for instance similarity between human and robot motion in pose configuration or end-effector position (see Sec. 5.4.1 for discussion of the correspondence problem), similarity in movement style, and choices regarding spatial orientation/location change. Choices have effects on the type of motion that can be realized on the robotic platform, as well as the user's perception of the motion.

For devising a mapping, the first choice that needs to be made regards whether it is more important to devise a mapping between end-effector positions (e.g., similarity between human hand location and the gripper of a robot arm) or a mapping to achieve pose similarity (as in our prototype). For human-humanoid imitation systems, the robot has a limited workspace. Imitation is convincing only in a particular human motion range. Consider also motion speed: if the robot has a lower speed of motion than human users, it will not be able to closely match a human's motion at high speeds. Dancers in the evaluation reported exploring how the robot's motion responded to theirs (D2, D3), including how the robot would respond to motion it cannot reproduce, such as jumping or lifting the shoulders. Moreover, they reported that the robot moves in a way that is not human-like, as it can isolate joints and it does not have a spine (D1). For the prototype described here, only in-place motion is considered, no location change. Locomotion and translation involve movements that result in a change of the robot's position in space (whereas configuration change consists of in-place body movements) [245]. Such motions could be considered for inclusion, but introduce complications for our prototype as a single RGBD sensor such as Kinect assumes a frontal position. The user is required to stay in the sensor's field of view in order to be sensed.

Orchestration

Regarding *Orchestration*, user attention is a factor of importance. Our prototype required changing the orientation of the body towards the Kinect, robot or the screen, visually attending to the screen and the robot, and ensuring body data can be captured with Kinect. Sensors that need to frontally capture human joint position data should be placed so that the user can orient themselves toward the robot. Mirroring of human motion by the robot (rather than imitating) while the human and the robot are facing each other was considered more familiar in the user evaluation (from viewing oneself in the mirror). If a user whose motion is imitated also needs to give commands using a GUI, the interaction design relating to the choice of input modes requires consideration, to avoid recording motion that should not be reproduced. For instance, during the evaluation, the user's leaning motion towards the computer to give start and stop commands was also reproduced. This should be avoided, for example by using alternative ways of giving commands to start and stop recording, e.g., with gesture control, speech commands, a physical button, or working with an additional person who starts and stops the recording. In the revised prototype, a countdown timer was used to allow the user to get into position prior to starting the recording.

The prototype required changing the orientation of the body (towards the Kinect or the screen), visually attending to parts of the system (the screen and the robot), and ensuring body data can be captured with Kinect. To minimize the need to continuously monitor all devices that are located in different places, the interaction could be focused on the robot, while an external screen is used for representing recorded motion, viewing recorded motion (e.g., in absence of a physical robot), and for setting the recording settings. To what extent an external screen is necessary depends on the chosen way of giving commands.

Integration

Decisions need to be made regarding which joints are mapped. P1 reflected on adding the possibility to the GUI to select individual limbs for recording. For *Integration*, one important consideration is that such systems make assumptions regarding bodies of users (see [256]) and their movement capabilities. In the prototype, assumptions arise due to the use of the Kinect and the mapping that is made from human to the robot, for example that users have two arms and are able to move them, which is not the case for everyone. Ideally, options should be given to customize which joints are detected and the robot's motion in response to human motion. But perhaps more importantly, other ways of programming the robot should remain available (e.g., providing code or specifying joint angle values).

Metaphor

Regarding *Metaphor*, dancers in the evaluation reported having the experience of embodying the limitations of the robot, or of the robot's limitations restricting their movement (D1). Some compared the interaction to the practice of being a movement teacher (teacher-pupil metaphor, D2), or reported experiencing the interaction as a conversation (D2, D3), communication or duet. A sense of connection or empathy was also reported.

Space-time mapping

Our prototype has a space-time to space-time mapping, from human movement in 3D space to robot movement in 3D space preserving timing. Generally, if there is both a physical robot platform and a virtual animation of a robot, this means that there are two space-time mappings in the system. Systems can deviate from each other. The program simulating the virtual robot can include a physics simulator, but that is not necessarily the case.

5.6 Conclusion

We developed a prototype for robot animation of the Pepper robot, which was revised after evaluation with dancers and programmers. The goal of the system is to facilitate robot animation by users who may not have sufficient programming skills but who do have other relevant expertise for developing contextually appropriate human-robot interactions. We extend a design space for computer animation so that it becomes applicable for describing robot animation systems that implement a human-humanoid pose-matching imitation system. We identify interaction design considerations connected to system architecture choices.

Acknowledgment

Supported by TU Wien Doctoral College TrustRobots, FWF #ConnectingMinds grant to the project Caring Robots // Robotic Care (CM 100-N), and a LINK-Masters project grant by Stiftung Niedersachsen to *“AI Tool for Dance and Choreography: It’s embodied and personalized, dare improvise?"*

CHAPTER 6

Human-In-The-Loop Error Detection in an Object Organization Task with a Social Robot

This dissertation chapter has previously been published as the following publication: Frijns, H.A., Hirschmanner, M., Sienkiewicz, B., Hönig, P., Indurkhyam B. and Vincze, M. (2024) Human-In-The-Loop Error Detection in an Object Organization Task with a Social Robot. Frontiers in Robotics and AI, Sec. Human-Robot Interaction, Volume 11. https://doi.org/10.3389/frobt.2024.1356827

6.1 Abstract

In human-robot collaboration, failures are bound to occur. A thorough understanding of potential errors is necessary so that robotic system designers can develop systems that remedy failure cases. In this work, we study failures that occur when participants interact with a working system and focus especially on errors in a robotic system's knowledge base of which the system is not aware. A human interaction partner can be part of the error detection process if they are given insight into the robot's knowledge and decision-making process. We investigate different communication modalities and the design of shared task representations in a joint human-robot object organization task. We conducted a user study (N=31) in which the participants showed a Pepper robot how to organize objects, and the robot communicated the learned object configuration to the participants by means of speech, visualization, or a combination of speech and visualization. The multimodal, combined condition was preferred by 23 participants, followed by 7 participants preferring the visualization. Based on the interviews, the errors that occurred, and the object configurations generated by the participants, we

conclude that participants tend to test the system's limitations by making the task more complex, which provokes errors. This trial-and-error behavior has a productive purpose and demonstrates that failures occur that arise from the combination of robot capabilities, the user's understanding and actions, and interaction in the environment. Moreover, it demonstrates that failure can have a productive purpose in establishing better user mental models of the technology.

6.2 Introduction

In Human-Robot Interaction (HRI) scenarios, failure situations invariably arise despite the best efforts of system designers. These failures can be caused by multiple factors, such as sensor noise or misinterpreted user input. A human interaction partner may be able to remedy errors. Existing work dealing with failure with the help of a human interaction partner often focuses on the communication of robot failures that the robotic system is assumed to be aware of. For example, in the works by [286, 60, 61], failure is modeled as a failure state during plan execution by a robot, for instance, getting stuck on a carpet while navigating [286], or being unable to pick up an object as it is located underneath another one [61]. However, there is a gap in the literature when it comes to designing robot communication that makes it possible for the user to spot errors that went undetected by the system itself.

Our general focus is on situations where there is an error in the system's knowledge base that the system is unaware of but that a user may notice. This requires that a human user is aware of the current state of knowledge of the robotic interaction partner. When the robot's knowledge is conveyed to a human user by means of a shared task representation, this representation can subsequently be inspected, verified, and corrected by the user, if necessary.

We consider the specific scenario of a robot at a user's home with the task of tidying up. It needs to know where each object is supposed to go, i.e., this user's personal preferences. When building a system for performing household tasks, it is necessary to consider how to represent the robot's knowledge to a human user and what types of failures can be expected to occur. In our study, a human shows a robot where to place certain household objects on a shelf.

Our work has two main aims. Our first aim is to find out how to support human-in-the-loop error detection in an object organization task with a robot through the communication of the robot's knowledge base. Our second aim is to investigate if and how the errors that occur with a functional system fit into the current understanding of failure in HRI scenarios.

Towards the first aim of supporting human-in-the-loop error detection, we developed representations of a robot's knowledge base for an object organization scenario. We conducted a user study in which the robot communicates a representation of its knowledge base (spatial locations of the organized object) to the user. In the study, we asked the participants to organize common household objects on a shelf. The robot then communicated its understanding to the user in one of three conditions; using 1) a visualization on its tablet, 2) speech, or 3) a multimodal condition combining both speech and visualization. See Figure 6.1 for an overview. We compare different output modalities, as it is not self-evident which modality will be preferred due to the spatial nature of the task involving an embodied robotic platform and objects organized at different spatial locations. Speech has an advantage, as it allows a user to focus their visual attention on the configuration of objects, while listening to the robot verbally indicating the objects' locations. A potential disadvantage is that speech may be perceived as slow, and thus, visual communication may be preferred. Combining the modalities may result in information overload and/or combine the disadvantages of both modalities. Hence, it is important to investigate this issue empirically. We analyzed the results to understand the advantages and disadvantages of different interaction modalities and errors that occur in such scenarios.

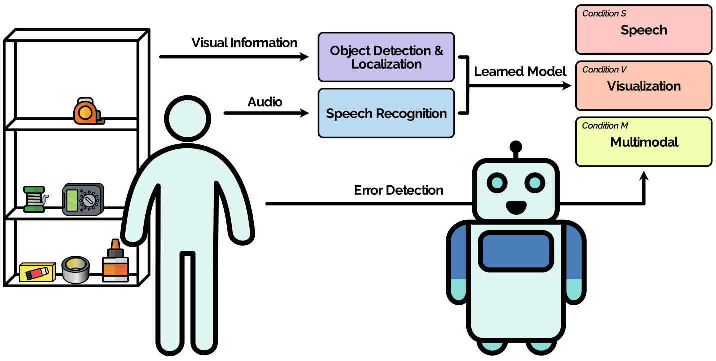


Figure 6.1: Overview of the system, setting, and study conditions. The task for the user is to arrange objects on a shelf to teach the robot the preferred configuration. The robot detects the position of the objects (Object Detection & Localization) and communicates the learned object configuration to the user. We compare three different types of output: Speech, Visualization, and a combination of speech and visualization (Multimodal). The user checks the output of the robot to detect errors (Error detection).

Our second aim is to investigate if the errors that occur in a study with a functional system fit current failure understanding in HRI. Often, pre-programmed, pre-determined or wizarded errors are used in studies on failure in HRI without studying other potential errors (see e.g., [157, 196, 115, 114, 189]). This means that the failures that occur in such experiments do not include potential user errors or errors that arise from interaction with the environment but rather focus on hypothesized robot failures. Working with a functional system that incorporates object detection allows for a more realistic investigation of errors that occur in service robotics (and in object organization tasks specifically) and to determine which representation design is adequate in a naturalistic task setup. In our study, we did not pre-plan the failure situations. We programmed the robot to go through the interaction script and perform an object detection routine fully automatically. As we used a functional object detection system in which participants were not constrained in the way they organized objects or interacted with the system, we were able to find out which types of errors actually occur. A subset of encountered failure cases arose from user curiosity regarding the capabilities of the system and do not fit into current failure taxonomies in HRI, which often assume a single source of failure. In the user study, we encountered failure cases that have a function in user learning. While these cases are failures in terms of not achieving the performance of the intended function or task, they do perform a productive purpose in terms of contributing to user learning or satisfying user curiosity. The cases we observed describe a type of failures that arises from a combination of user actions on the environment, user expectations, and robot capabilities/limitations, which extends current understanding of failure in HRI.

In line with the aims sketched above, our work is guided by the following research questions:

RQ1 *How to support human-in-the-loop error detection in an object organization task with a robot?*

RQ2 *How do the failures that occur with a functional system fit into current understanding of failure in HRI?*

The contributions of this paper are as follows:

1. We present proposals for the design of representations of robot knowledge of object arrangements to human users by means of speech and/or visualization (Section 6.4.3, 6.4.4), towards *RQ1*;
2. We present findings from a user study regarding user preferences on communication of robot knowledge in a human-in-the-loop error detection task (Section 6.6, Section 6.7.1), towards *RQ1*;
3. We introduce the concept of productive failure in HRI, which is not part of any existing HRI failure taxonomies (Section 6.7.2), towards *RQ2*;
4. We argue for an understanding of failure in HRI as an interconnected phenomenon that develops over time and that can involve humans, robots, other agents, objects acting in the environment, which extends the understanding of failure in HRI failure taxonomies beyond one that considers failures as having a single source of origin (Section 6.7.2), towards *RQ2*.

6.3 Related Work

In this section, we discuss related work on multimodal and transparent interfaces for robotic systems in Section 6.3.1, in line with our first aim. In Section 6.3.2, we outline concepts of failure and failure taxonomies in HRI, in line with our second research objective.

6.3.1 Design of Multimodal and Transparent Interfaces in HRI

Previous research has investigated the use of individual modalities in HRI scenarios, such as displaying facial expressions for giving feedback [190], comparing unimodal displays of emotion [281], or investigating preferences regarding a person-following robot's auditory feedback behavior [200]. In a Human-Computer Interaction context, multimodal input has been argued to offer advantages such as supporting user preferences and user learning, and reducing cognitive load and user errors when fusing information from user input modes [73, 109]. Advantages of multiple output modalities for communication from system to the user may be providing information in complementary forms [207] or improved inferring of past causal information [116]. In HRI, there is a gap regarding research comparing multimodal with unimodal inputs/outputs. In our work, we compare user preferences regarding the use of visual and auditory modalities, and a combination of the two, for presenting study participants with the configuration of objects detected by a robot.

Research on information presentation in relation to object configurations in HRI includes work on disambiguation or grounding of human requests, for instance for the PR2 robot [111]. [249] argue that the use of natural language can result in ambiguous requests, and propose that using visualization can aid with disambiguation. In their user study, participants verbally describe an object, and the system visually indicates the inferred object by means of a head-mounted device, projector, or screen. [69] propose the Grad-CAM RGB method, which aims to identify specific objects in RGB-D images for human-robot collaboration in unstructured environments. They draw bounding boxes to indicate image regions in response to textual queries. [61] investigate semantically descriptive explanations of objects in a scene to help end users identify reasons for robot failure to pick up an object.

Such information presentation of a robot's perception and processing is a form of transparency. Transparency refers to information being provided by a system with the aim to give a human end user a better understanding of what the system is doing and why [100]. This can concern information regarding a robot's internal processes, for example, inferred commands [211], robot plan representation [301], learned words for objects and verbs [125], or providing a rationale for robot decision-making [196]. Hence, human interaction partners can better assess the robot's state and form more accurate expectations regarding its behavior. Several works have investigated the design of transparent interfaces in HRI. [294] developed an interface that visualizes the decision-making process of a robotic system and allows for inspection and editing of the knowledge graph. They argue that such an interface can support the sensemaking of robot decision-making. [211] performed a user study in which they compared a baseline of transparency through speech, pointing and gaze to a combination of the baseline with visualization-based transparency on screen, and to a combination of the baseline with visualization in Virtual Reality. The scenario concerns object manipulation and surface cleaning by the PR2 robot in response to human commands. Visual transparency enhanced the accuracy of commands and required less time, and increased the number of observed pointing gestures by participants. The screen-based visualization condition was preferred by participants. [125] investigated a language-learning scenario of a human tutor teaching a Pepper robot words for objects and actions. They compared different transparency strategies, namely pointing and gaze by the robot to request information regarding particular objects, and visualization of learned words on Pepper's tablet. The visualization led to a higher self-assessed knowledge of the system state as compared to pointing. Several studies looked at the connection between errors and transparency, notably also in connection to trust [196, 114]. [115] ran a user study in which the BERT2 robot carried out a kitchen task, either with or without an error, and with or without verbal communication when correcting the error. They found support for their hypotheses that higher transparency mitigates dissatisfaction in case of errors, and that participants prefer a more communicative system.

In contrast to works that investigate how to communicate robot *failure* [61, 60], we investigate how to communicate a robot's *knowledge base* so that the user can inspect whether it is correct or a failure has in fact occurred. This is similar to work on transparency, e.g., by [125, 211] who visualize recognized speech, objects, and actions and compare different ways of communicating. However, in our work, we explicitly focus on human-in-the-loop error detection.

6.3.2 Failure and Errors in HRI: Concepts and Taxonomies

In the definition by [35], a failure occurs at the level of the task or functionality that a robotic system is supposed to perform: *"A "failure" refers to a degraded state of ability which causes the behavior or service being performed by the system to deviate from the ideal, normal, or correct functionality"* [35, p. 9]. This definition of failure is used for the taxonomies by [309, 134]. Similarly, failure has been defined as *"an event that occurs when the delivered service deviates from correct service"* [258, p. 345] or the *"inability of the robot or the equipment used with the robot to function normally"* [38, p. 423]. In the works by [286, 60, 61], failures are conceptualized as actions in a plan that fail, which results in a failure or halting of the robotic system's plan. Several authors describe a relation between failures, errors and faults. According to [35, 38, 134], faults may lead to errors, which in turn may lead to failures. Applying this description to the task in our study, we can describe an object that is missing from the knowledge base representation as a failure in the robot's task to detect organized objects and convey its knowledge base to the human interaction partner. This can occur due to an object detection error (e.g., white glue bottle is not detected), which can be caused by the fault of overexposure of the camera image.

HRI failure taxonomies are categorizations of failures that occur in HRI scenarios. Although it is unlikely that all possible robot failures can be identified for mobile robots in changing environments combined with a wide variety of possible interactions [134], several authors have proposed failure taxonomies that aim to classify failures. See Table 6.1 for an overview of failure taxonomies in the HRI literature. [134, 38] classify based on the source of failure. The taxonomy by [134] distinguishes between technical and interaction failures. Human errors are a subcategory of interaction failures, and can be mistakes, slips, lapses, or deliberate violations, based on the work by [223]. [38] propose a taxonomy of failures for mobile ground robots (Unmanned Ground Vehicles, UGVs). They categorize failures according to their source, and failures are divided into physical and human failures. [279] classify trust-relevant failures in HRI based on a different choice, namely if the action of breaking trust was by the system or the user, and if this agent was supposed to act this way. They identify the failure types of Design, System, Expectation and User.

Some authors argue that failure categorizations should be more human-centered. [277] write that the majority of HRI research takes a robot-centric perspective on failure. They note that failure is not only based on robot capabilities but also on its alignment with the context and whether the robot's behavior is socially aware. [309] argue that while existing typologies classify failures according to what technically caused the failure, users do not know what caused the failure on a technical level, but make an assessment based on the information they have. They propose an output-oriented typology of performance failures, classifying them into logic, semantic and syntax failures. They classify performance failures in HRI based on expected versus actual output. In their taxonomy of social errors in HRI, [278] classify according to *"five categories that humans adopt to perceive socio-affective competence and social relationships"* [278, p.14] (see Table 6.2).

Existing failure taxonomies and conceptions of what failure entails influence study designs in HRI. For example, [157, 158] performed Wizard-of-Oz studies in which they based the failure types that a robot made on the failure taxonomy by [134]. This illustrates that the way failure is conceptualized in taxonomies and definitions is important for the studies that will be conducted and contribute to our knowledge of the topic of failure in HRI, as well as the mitigation strategies that are proposed for robotic systems that encounter failure situations. In our work, we investigate how the errors and failures we encountered in our user study fit in current understandings of failure in the HRI literature, and argue how this understanding should be extended. In contrast to existing taxonomies that classify failures based on their source, we argue that HRI understanding of failure should acknowledge failure cases that arise due to a combination of technical capabilities (and limitations), human understanding and behavior, and interaction with the environment and other agents and objects in the context. Moreover, in our study, we observe user actions that do not fit neatly into the way “human errors" are currently described in such taxonomies.

6.4 Designing a System that Conveys Detected Objects to a User

Our goal was to compare different ways of communicating robot knowledge of object configurations. In the envisioned scenario, a study participant organizes objects at a particular location. The robot learns user preferences from observations. The robot stores the object locations and conveys the configurations back to the user, who checks if these are correct, so the robot can (hypothetically) organize the objects by itself later. We investigated different ways of communicating the stored object locations. The communication of the stored object locations thus functions as a shared task representation. The concept of shared task representations has been proposed as a way to reduce common ground uncertainty in human-robot joint activities [51, 100]. Common ground [46] refers to a set of mutual or shared beliefs held by the human(s) and robot(s) involved in the interaction. We investigated how to design shared task representations and which interaction modalities are suitable for human-in-the-loop error detection.

6.4.1 Object Detection

For object detection, we integrated a YOLOv5 [146] implementation with ROS Noetic. The model was trained to recognize 11 objects (tools and objects commonly found at a workshop, e.g., measuring tape and wood glue). We used the copy-paste data augmentation technique to generate a diverse dataset by pasting images on random backgrounds [103]. We recorded 100 images per object from multiple viewpoints. We applied the DINO-ViT algorithm [13] to generate object masks. After filtering corrupted masks due to reflections, 20-50 image-mask pairs were kept per object, which was sufficient to cover the full surface of each object. The masked objects were randomly pasted onto backgrounds from the COCO dataset [175], resulting in 50,000 images with more than 3000 instances per class. YOLOv5 was trained using the parameters set as suggested by the original implementation of [146]. From the 11 objects on which the detector was trained, we used the 6 objects with the most reliable object detection performance in the specific setup of the experimental study. We used OpenCV’s ArUco marker detection [203] to detect the location of the shelf to determine the object positions in relation to it. Object locations were detected in 2D image space.

See Figure 6.2 for an example of the object detection running on the camera stream from one of Pepper’s cameras. During the experiment, the area in which the robot had to detect objects in the cupboard required the robot to move its head up and down to cover the area. The images from the camera stream had some motion blur. We combined the detection results from multiple camera images that covered the detection area, which took approximately 10 seconds. To give an idea of the system's task performance, in the experiment there were 122 runs of the object detection routine, and in each of these runs 6 objects should be detected (732 in total). Excluding the 6 cases in which an object was fully hidden from view, we had 63 cases in which an object went undetected, yielding a performance of $(1-63/726)*100\% \approx 91.3\%$ on the experiment task. The coil was detected 93.4% of the runs, the wood glue for 87.7%, the glue 84.4%, the tape 95.9%, the measuring tape 86.9% and the multimeter 100%. Note that in some of these cases, objects were partially hidden behind other objects. See Section 6.6.3 for more details.

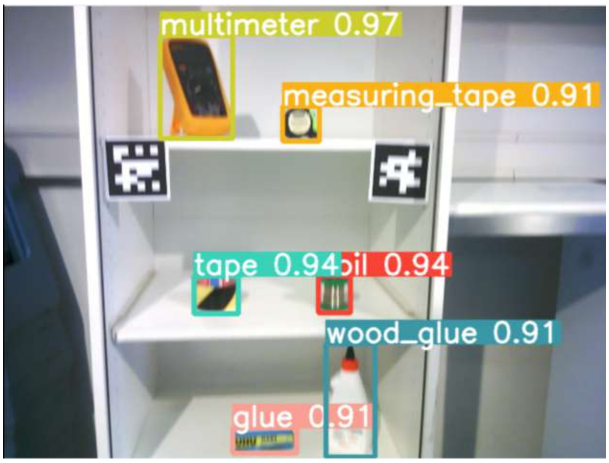


Figure 6.2: Example of the object detection using an inbuilt camera on the Pepper robot.

6.4.2 Robot Behavior

We programmed a Pepper robot to speak, perform gestures, and display information on its tablet with the Python SDK with NAOqi [255]. The robot was programmed to execute slight motions to indicate that it is active, perform co-speech gesture and look at the participant (using face detection), which was implemented with the NAOqi API modules *ALBasicAwareness* and *ALAnimatedSpeech*. During object detection, the robot oriented itself to face the cupboard and moved its head up and down, verbally indicating that it was scanning the area. This behavior was intended to be both functional and communicative; the robot's motion was used to collect image data of the full view of the cupboard and also expressed that object detection is being performed.

When the visualization was used, the robot turned to face the participant so that the left side of the tablet visually matched the left side of the cupboard from the participant's perspective. This way, the participant did not need to cross-compare the visualization to the object configuration in the cupboard, see Figure 6.3.

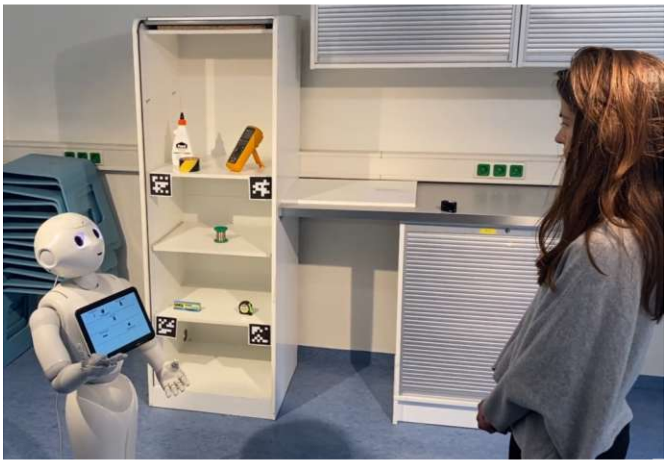


Figure 6.3: Robot turns to face the participant in conditions with visualization.

6.4.3 Speech Modality

The spoken description of object configurations by Pepper was based on research on descriptions of such configurations by native German speakers [267, 268]. Speakers tend to look for a salient object (a *relatum*), in relation to which they can describe the location of the target object. If a salient object is not available, they may begin with a reference region instead. Descriptions are usually sequential, following a continuous trajectory [267, 268]. We designed the speech condition as follows. We used two prepositions from [111], namely *"to the left of"* and *"to the right of"*. A description was generated for each object. This is done on a per-area basis (e.g., considering those objects that are on a single shelf). When generating a description, the leftmost object was taken first and described using the approximate location of the object relative to the area, e.g., *"on the left of"*, *"somewhat to the left of"*, *"in the middle of"*, *"somewhat to the right of"*, *"on the right of"* + area name. Each next object in the area was described by referencing the previously described object plus the approximate location in the area. This generated descriptions such as: *"The wood glue is on the left of the top shelf. Next to the wood glue is the measuring tape, which is in the middle of the top shelf."*

6.4.4 Visualization Modality

Visualization options that were considered include an abstracted visualization with icons and text, a knowledge graph representation that closely matches the robot's internal representation (as in, e.g., [294]), or a camera stream with a visual overlay (e.g., [211]). In our task scenario, a camera stream representation would require either a static composition of multiple images from the video stream (which would result in a cluttered view that does not fit on the tablet in a way that individual objects can be distinguished); or it would need to change dynamically (which we expected would make it more difficult to keep track of the knowledge base). However, such a representation can be relevant for a follow-up study that involves troubleshooting when participants want to know why a particular object is wrongly detected or not detected. An internal graph representation was considered, but expected to result in a cluttered view as well. Therefore, we chose to represent the scene with an abstracted visualization, in which icons with text represent the objects. This representation provides an uncluttered representation of the scene that fits the tablet dimensions, as shown in Figure 6.4.

We used a JavaScript library [172] in a Choregraphe application package for the visualization, which could be dynamically adapted using Python commands. The visualization consisted of icons and names of the detected objects. Objects were displayed at a particular location, assigned to an area in a similar way as with speech (Section 6.4.3).

6.5 Study: Human-In-The-Loop Error Detection

A within-subjects study was conducted in which participants interacted with a Pepper robot in three conditions. The participant placed objects on a shelf to "teach" the robot their desired locations. After the objects were placed, the robot either communicated the object locations verbally (condition ***Speech (S)***), by means of a visualization on its tablet (condition ***Visualization (V)***, see Figure 6.4), or using both visualization and speech (condition ***Multimodal (M)***). The communication of object configurations was done across communication channels that were additional to the robot's nonverbal functional behavior (performing the object detection routine, Section 6.4.2), interactive behavior and verbal instructions to the participant.

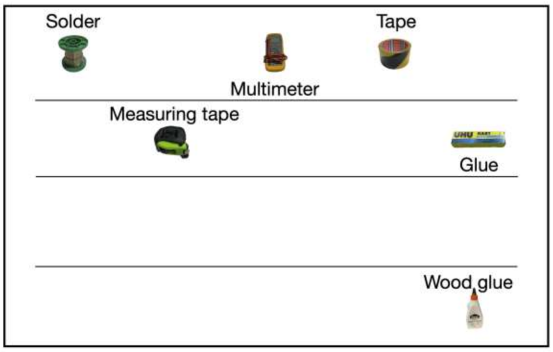


Figure 6.4: Example visualization in the Visualization condition.

We investigated the issue of human-in-the-loop error detection from different angles, namely by considering participant behavior, participant interpretation of the system, failures that occurred with a functional object detection system, and participant preferences. The study research questions were as follows:

RQ-Participant preference: Which condition do participants prefer? Why?

RQ-Task Load: Do participants perceive a different workload between conditions?

RQ-Error: Do participants notice all errors? Is there a difference between conditions?

RQ-Mental model: How do participants construe the way the system works?

RQ-Participant behavior: Do participants use strategies to test the system, and when do they do so?

6.5.1 Study Protocol

The experiment took place in a museum of science and technology in Vienna, Austria, in a separate room from the exhibition space. The interaction with the robot, questionnaires, and study procedure were in German. The study procedure and informed consent process were peer reviewed by the Ethics Committee at our university. The protocol included an explanation of the study, informed consent, interaction with the system in three conditions including completion of a survey form, and an exit interview.

First, participants were informed regarding the purpose of the study, data collection and storage, that they could opt out of the study at any time, and they were given researchers' contact information. This information was included on informed consent and data consent forms that participants were asked to sign. Participants were asked to complete a short personal information questionnaire, and were introduced to the robot, how it worked, and its different sensors (Figure 6.5). The robot was programmed to indicate its sensors and describe their function, which was then repeated and explained by one of the researchers. This introduced the participant to the robot's movement and enabled us to check that the robot's speech volume was adequate for the participant. The researcher answered questions and explained the task.

In each interaction condition, the task for the participant was to place several objects on the shelves of a cupboard. First, the robot greeted the participant and announced it would scan the area. It then "scanned" the area by moving its head up and then down. Then, the robot asked the participant to arrange the objects in the cupboard and to indicate when the task was complete. The participant put the objects in the cupboard (wood glue, tape, measuring tape, a multimeter, a glue package, soldering tin) (Figure 6.5). When the participant said they were done (when the speech recognition detected the word *"fertig"/done*), the robot asked the participant to step aside. The robot performed the scanning motion again, during which object detection was running on the video stream of one of Pepper's cameras. Depending on the condition, the robot spoke out the object locations (*Speech*), displayed them on the tablet (*Visualization*), or conveyed them by both speech and visualization (*Multimodal*). The robot asked the participant to confirm whether it was correct. If the participant confirmed, the robot thanked the participant, concluding the interaction. If the participant indicated it was incorrect, the robot announced it would scan the scene again from another perspective. It moved to the right and scanned the scene again, conveyed the object locations, and again asked the participant to confirm. If still incorrect, the robot stated it could not resolve the error. In either case, the robot thanked the participant, concluding the interaction. After each condition, participants completed a questionnaire. After interacting in all three conditions, participants were asked if they had time for an interview and were explained how the system worked.

The interaction with the robot was programmed to function fully automatically. As it was not the focus of our study, we chose to trigger the speech detection remotely if the built-in speech recognition did not function after the first two tries by the participant.

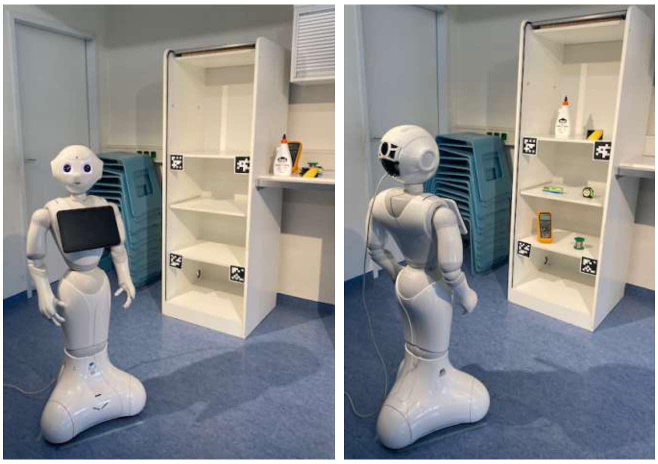


Figure 6.5: During the explanation (left); after the participant arranged objects (right).

The interaction of participants with the robot was video and audio recorded. While the robot was performing object detection, a recording was made of its camera stream. An interaction log was kept automatically. Prior to the experiment, participants completed a questionnaire on personal information. After every condition, participants completed a questionnaire asking if the robot made mistakes regarding the locations of objects, and if yes, which. Other questions on the survey included if the robot made any other mistakes, how the robot communicated the information, and Raw NASA-TLX [118] translated to German [90], to measure the users' task load, which contains subscales for mental, physical and temporal demand, performance, effort, and frustration. After each participant interacted with the robot in all three conditions, they completed a final questionnaire on their preferred condition and the perceived condition order as a manipulation check. Questions in the structured exit interview included *Which version did you prefer and why? How did the robot learn the locations of the objects?* See Table 6.3.

6.5.2 Pilot Study

A pilot study was conducted at a museum of science and technology (N=4, 1 woman, 3 men, age M=35.25, SD=6.02). The protocol as described in Section 6.5.1 was followed. One participant preferred condition *Multimodal*, stating it was the clearest, while three preferred *Speech* due to either not being able to see well on the screen, a concern that the visualization may not be clear enough in case the objects would be organized in a chaotic way, or feeling stressed with the combination and not always trusting the way it was displayed. After the pilot, minor changes were made to the questionnaire and the robot behavior. The robot's behavior was changed so that it introduced itself and its sensors (instead of a researcher introducing the robot), which also allowed for checking if the robot's speech volume was set at an adequate level for the participant prior to interaction in an experimental condition. The questionnaire was changed to ask the participant if the robot made an error in two different ways instead of once. After the revision, it asked whether the robot made an error both in terms of the object positions or if it made any other error.

6.5.3 Participants

For the main study, 33 participants were recruited in a museum of science and technology. Conditions were counterbalanced for 33 participants; the order of execution of counterbalanced conditions was randomized across all participants. This meant that 5 participants interacted in each of the condition orders VMS, SMV, and MSV, and 6 in each of the condition orders VSM, MVS, and SVM. Two of them did not pass the manipulation check: they did not correctly identify how the robot communicated the object positions on the survey forms after each interaction, nor on the final survey form.

In Section 6.6, we report on the data of the remaining 31 participants (16 men, 15 women). Their ages (M = 39.1 SD = 15.87) ranged from 17 (with parental consent) to 76. Nine participants indicated they had a robot at home (vacuum cleaner robot or LEGO). Nineteen participants stated having seen a robot, nine of whom had interacted with a robot, and two had programmed a robot (LEGO Mindstorms). Self-assessed computer programming experience was rated by 18 participants to be 1 (very inexperienced, i.e., have not programmed anything before), by 5 participants to be 2, by 1 participant to be 3, by 5 participants to be 4, and by 2 participants to be 5 (very experienced, i.e., professional programming knowledge).

6.6 Results

Thirty-one participants interacted with the system in all three conditions, yielding 93 interactions. The object detection routine would run once or twice per interaction, depending on the participants' response to Pepper's question whether the representation was correct after the first run of the object detection routine; if not, this routine would be performed again. This resulted in 122 runs of the object detection routine. Some interviews could not be used due to equipment failure. We decided to keep these participants in the analysis of the main study, since the equipment failure did not interfere with the interaction or the questionnaires. Additionally, the interview was designed to be voluntary (and presented as such to participants) to give participants the opportunity to clarify their questionnaire responses. The interview responses of twenty-four participants were transcribed and analyzed by coding the responses to the individual interview questions on preferred condition, the way the system works, and the way participants organized objects.

The interview analysis was conducted as follows. The first and the second author first familiarized themselves individually with the data, proposed a coding scheme, discussed the coding schemes together and then agreed on a coding scheme. See Table 6.4 for an overview of the final coding scheme. Each of them individually coded the responses of the participants to the interview questions 2-5. For 89.7% of the 194 assigned codes the coders were in agreement (meaning that the two coders assigned the same code to the response). Coding differences were resolved by discussing each of the differences to arrive at a final joint coding.

6.6.1 Participant Preference for Interaction Modality

To answer **RQ-Participant preference**, we looked at the questionnaire results and interview responses regarding the preferred condition. On the final questionnaire, 23 participants preferred condition *Multimodal*, 7 preferred condition *Visualization*, and one preferred condition *Speech*. A multinomial test [230, p.142] yielded a p value of $p = 2.315 * 10^{-6}$, thus we should reject the null hypothesis that each condition is preferred equally. An exploratory data analysis showed a difference in self-reported computer programming experience for participants who preferred the condition *Multimodal* (M = 1.57, SD = 1.16) and for participants who preferred the condition *Visualization* (M = 3.29, SD = 1.38). We performed multinomial logistic regression to check if “Programming Knowledge” can be used as a predictor variable for the preferred condition. The likelihood ratio test resulted in a $p < 0.05$ indicating that the model including programming knowledge as the predictor variable is significantly better than the null model. The coefficient for “Programming Experience” was 0.9052 (std err$= 0.353, z = 2.566, p < 0.05$), indicating that for every unit increase in “Programming Experience”, the log likelihood of participants preferring *Visualization* as compared to *Multimodal* increases by 0.9052. In a model comparing preference for *Visualization* to preference for condition *Multimodal* that included gender and age besides programming experience, the programming experience was the only predictor variable with $p < 0.05$.

In the interviews, participants stated a variety of reasons for preferring a specific condition. Reasons for preferring *Multimodal* included that it was a double confirmation regarding what the robot saw, it being clearest, having to make more of an effort to think along with speech-only or the risk of forgetting or mishearing with speech-only, being better able to compare it, the participant having a left-right weakness so seeing makes it easier, it reducing the chance of making mistakes, and the combination resulting in more humanlike communication. Reasons for preferring *Visualization* included that it is faster, being used to tablets, finding it easier to see rather than hear, and it being more effort to check if both visualization and speech were correct. A reason for preferring *Speech* included that seeing it on the tablet was too much information.

6.6.2 Task Load

The range for the 6 NASA-TLX subscales are from 0 (low) to 100 (high), and so are the total NASA-TLX scores that are calculated here by summing the subscale scores and dividing by 6. The total NASA-TLX means per condition (calculated using the scores of all participants) were as follows. For *Visualization*, the score was M=12.69 (SD = 8.00), for *Speech* M=14.41 (SD = 10.46) and for *Multimodal* M=10.78 (SD = 6.43). The ratings were rather low for all three conditions; we observed a floor effect that makes it difficult to differentiate between groups. This tendency was not observed in the pilot study. See Figure 6.6 for an overview of subscale ratings averaged across all participants.

We performed Friedman tests [230, p. 154] with Python [271] for the total NASA-TLX scores and the subscales. We used the Friedman test, as the data is ordinal, and it is a repeated-measure (the study is within-subjects). We did a Bonferroni correction on the α level $(0.05/7 = 0.007)$. The Friedman test for Frustration yielded a p-value of 0.0029<0.007. The other tests were not significant.

The mean scores for Frustration were as follows: for the *Visualization* condition M = 15.32 (SD = 18.21), for the *Speech* condition M=14.68 (SD = 13.72), and for the *Both* condition M=9.52 (SD = 9.07). As the Friedman test yielded a significant difference in Frustration among the different conditions, we did three Wilcoxon signed-rank pairwise comparisons for Frustration (following [230, p.155]), applying a Bonferroni correction: $0.05/3 = 0.01667$. We found no significant difference between *Visualization* and *Speech* $(Z = 52.0, p = 0.6456)$. We found significant differences in the

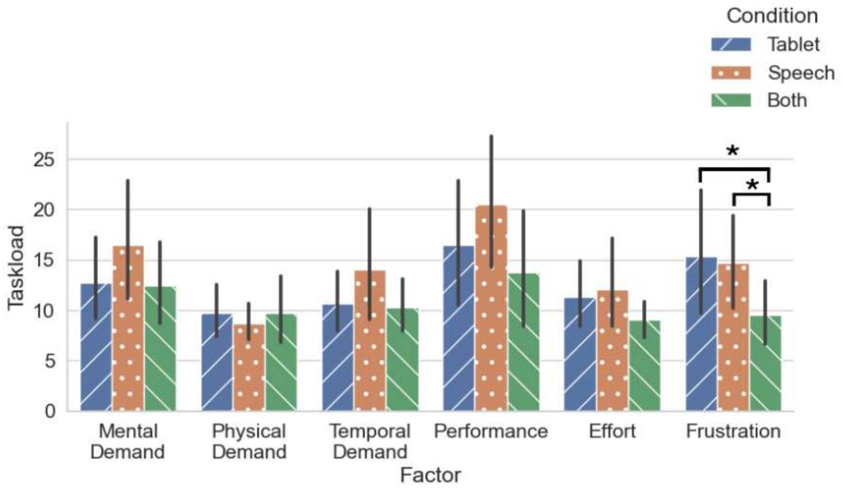


Figure 6.6: NASA-TLX subscale score averages of all participants (N=31). *p $<$ 0.01667

comparison of the *Speech* and *Multimodal* conditions $(Z = 2.0, p = 0.0035 < 0.01667)$ and the *Visualization* and *Multimodal* conditions $(Z = 3.5, p = 0.0030 < 0.01667)$. We can use $r = Z/\sqrt{N}$ to calculate the effect size for non-parametric tests [230], which gives us a medium effect (0.36) for the comparison between the *Speech* and *Multimodal* condition and a large effect (0.63) for the comparison between the *Visualization* and the *Multimodal* condition.

To get an indication whether this may be due to one condition having more perceived error occurrences, we performed two Cochran's Q tests (using [210]). We performed a Cochran's Q test to determine if there were differences between conditions whether participants reported errors or not, and a Cochran's Q test if there were differences between conditions regarding whether there were any object detection errors or not. Interactions by participants with the system in which errors occurred were coded with a 0, and cases in which no error occurred were coded with a 1. We found no differences between conditions regarding the (perceived) occurrence of errors.

To answer **RQ-Task Load**, we found no significant difference between the total NASA-TLX scores. However, we did find differences for Frustration between the *Multimodal* condition and the *Speech* condition, and between the *Multimodal* and the *Visualization* condition.

6.6.3 Errors

To answer **RQ-Error**, we investigated whether participants noticed all errors and if this depended on the condition. This required identifying which errors occurred and comparing them to the errors that were reported by participants. We observed that participants were able to identify most errors in all conditions. In a few cases, participants did not notice errors, but these cases were too few to distinguish between conditions. We identify some modality-specific limitations. Further discussion is included in Section 6.7.1.

From the 93 interactions, in 41 cases the participant correctly identified that no errors occurred, while in 31 cases the participant correctly identified the errors that occurred. In 7 of those 31 cases, the participant responded to the robot that the representation was correct, but afterwards stated on the questionnaire that there was an error. In 12 cases, the participant stated there is an error but there was no object detection error per se (in those cases, only errors such as speech recognition errors occurred). In 9 cases, errors, inaccuracies or mismatches in the participant's judgement were observed (Section 6.6.3).

Object Detection Errors

We identified object detection errors (e.g., missing objects) by comparing the camera stream during object detection to the interaction log, which contained the results of the object detection. This was analyzed for each of the 122 times the object detection routine was performed. In 58 of these instances, one or more errors occurred. The main error concerned objects not being detected (69x), followed by detection of objects that were not there: in 7 cases, objects were detected that were in the data set but not in the set that participants were asked to organize (5x a marker, 2x pliers). Four object location errors occurred. The number of participants for whom one or more errors occurred was 11 (out of 31) during the first interaction, 11 for the second interaction, and 17 for the third and last interaction.

Errors According to the Study Participants

Participants were asked to indicate on the survey form if the robot made (1) an error regarding the object locations or (2) any other errors, and if so, which ones, see Table 6.5. In addition to the object detection errors described above (Section 6.6.3), participants also reported speech recognition errors and social interaction errors (e.g., not turning towards the participant while speaking, long response times).

Errors and Inaccuracies in Participant Judgement

In 9 of 93 cases there was an error or inaccuracy in participant judgement based on the survey form results and their response to the robot (5x *Speech*, 2x *Visualization*, 2x *Multimodal*). However, participants were observed to make statements towards the researcher or display nonverbal behavior that is indicative of them noticing an error in some of these cases. We list the specific cases below:

- Twice (in the *Speech* condition) objects were not detected but the participant said there was no error (to the robot and on the form). In one case, three objects were missing. In the video footage, this participant addressed the researcher in a questioning way, saying that objects were missing, but that other objects the robot talked about were correct.
- One participant indicated twice (in the *Speech* and *Multimodal* conditions) that there was an error and indicated one missing object, but an additional object was also missing.
- Twice, an additional "unknown" object was detected that was not part of the objects participants were asked to organize, but that was part of the dataset (pliers, marker), without the participant

indicating on the form or to the robot that there was an error. However, on the video data, one participant (*Visualization*) verbally remarked to the researcher that there was a marker, and one participant (*Speech*) initially nodded in response to each of the robot's statements about object positions, but stopped nodding once the robot mentioned pliers, which could indicate that the participant noticed something was wrong.

- Twice (*Multimodal, Visualization*), a participant noted that the robot could not indicate that objects were placed behind each other, but in principle no object detection error occurred. In the *Multimodal* case, one object was missing from the representation during the second scan, which was not remarked on by the participant on the survey form; the participant remarked that the robot was not able to recognize that objects were placed behind each other or on top of each other.
- Once in the *Speech* condition, a participant stated that there was an error without specifying, but no error could be identified by us.

6.6.4 Mental Model: Participant Interpretation of the Object Detection System

To answer **RQ-Mental model**, we coded the interview responses to the question *"How did the robot learn the positions of the objects?"* The participants made reference to scanning (12 participants), cameras or photography (7 participants), the objects being pre-programmed or the robot having an existing representation of the objects (6), the markers (5), seeing or eyes (3), the shape of the objects (3), with the text on the object (2), sensors (2), programming (1), the size of the object (1), the ultrasound and infrared sensors (1), or that it was done by a human (1).

In other words, participants explained how the robot detected objects by referring to the words used by the robot, to its behavior, what they could observe in the space, on the objects and in relation to the robot, and some made reference to its humanlike capabilities like seeing and reading. Participants had multiple sources of information available to them about the way the system worked: what the researchers told them at the start of the experiment regarding the robot's sensors and the task, the words and behavior of the robot, what could be observed by the participants (e.g., ArUco markers on the cupboard), and the events during the trials. For example, the mention of scanning can be connected to the robot stating it was "scanning the area". Some of the theories participants constructed were (partially) incorrect. For example, one participant stated: *"It obviously has a database of those objects, and it looks for those exact objects every time where they are. Last time I forgot the soldering tin outside. Then he didn't realize that the soldering tin wasn't there at all. So, obviously, he looks for exactly these objects every time on the shelf. And if one of them is not quite there, i.e., it's not on the shelf, but it's on the side, then he recognizes it as being on the shelf. Even if it is not on the shelf."* This illustrates how accidental events during the trial influenced the participant's understanding of the system. There was a technical failure during the trial: the soldering tin was detected as being on the shelf because the shelf area had been defined to be a bit wider than it actually was. So the soldering tin, which was forgotten by the participant and left next to one of the shelves, was also assigned to being on the shelf. This resulted in inaccurate understanding of the participant; the robot does not actively look for those exact objects.

6.6.5 Participant Behavior: Object Arrangements

To answer **RQ-Participant behavior**, we analyzed both participants' object organization strategies as well as their answers to the interview question that asked if they had a specific reason for the way they organized the objects. The strategies that were mentioned by participants in the interviews were coded, as were the strategies that were observed with regard to the object configurations.

In the interviews, several strategies were mentioned. Eleven participants mentioned testing the system in some way, for instance, by placing an object inside or behind another object, or rotating an object, to see if the system would still be able to detect the configuration. Ten participants mentioned organizing objects without any specific intention in mind. Five reported ordering objects by category (e.g., grouping adhesives or measurement tools). One other mentioned reason was making a clear, symmetrical arrangement. One participant stated: *"I thought I would give him easier tasks at the beginning and then more difficult ones. That's just the way it is with a child. You just increase it."* Another: *"So the first time I thought, nice and symmetrical, clear. The second time I made it a little harder, I think for him, because then I also put two things next to each other. And the third time, I turned the tube of glue on its side because I thought that if he only had the [brand name] as a pattern, he might stumble, and he did. It was mean, wasn't it?"*

For analysis of the object organization strategies, the first and third author agreed on a classification and then coded screenshots of the organized cupboard (inter-rater agreement 92.1%, mismatches were resolved jointly). We distinguished several strategies that participants may have thought would make it more difficult for the robot to detect objects, which were also reflected on during the interviews. See Figure 6.7 for an example. The number of times this happened increased from the first interaction. Strategies identified included stacking an object on top of another, placing more than three objects on a shelf, rotating objects, hiding objects from the robot's perspective or placing them in a way that they overlapped one another, and placing objects at different depths on the shelf. See Figure 6.8.

In the first interaction, an object was hidden behind another object zero times (0x), in the second interaction 4x, and in the third interaction 5x. Objects were placed in a way that one object overlapped another from the robot's camera perspective 3x in the first interaction, 6x in the second interaction, and 10x in the third interaction. Placing more than three objects on a single shelf occurred 0x for the first interaction, 3x for the second, and 5x for the third. Stacking of objects only occurred two times, both in the third interaction. For rotation, the coding required distinguishing between canonical and non-canonical views, where canonical views would be those views that people find more typical or easier to recognize [212]. Rotation of objects (a non-canonical view) remained more or less similar across interactions, with an object being rotated 12x for the first interaction, 8x for the second, and 16x for the third interaction (excluding measuring tape as it lacks a clear orientation).

6.7 Discussion

In this section, we discuss our findings regarding human-in-the-loop error detection to answer *RQ1* (Section 6.7.1) and how the observed failure cases align with current HRI understanding of failure to answer *RQ2* (Section 6.7.2) in line with the research objectives stated in the Introduction.

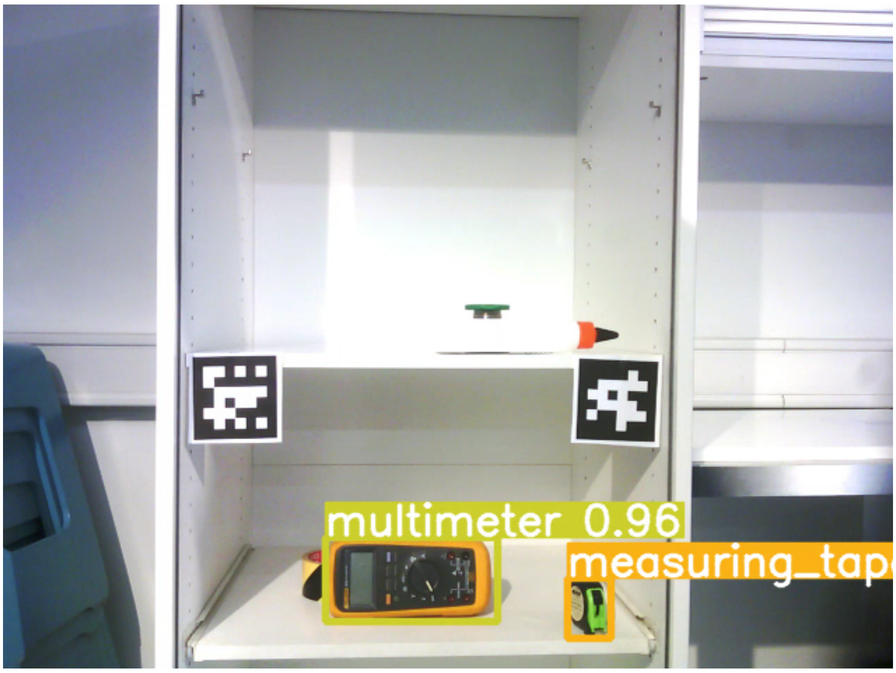


Figure 6.7: Example of an object configuration by a participant in which the objects are more difficult to detect by the system.

6.7.1 Supporting Human-In-The-Loop Error Detection

In this section, we discuss how to support human-in-the-loop error detection in an object organization task with a robot. Most participants (71.9 %) preferred to get the representation of the robot's knowledge both as a visualization and through speech, as this was a double confirmation. Moreover, the NASA-TLX subscale score for *Frustration* was lower in the *Multimodal* condition as compared to the single-modality conditions. As participants correctly identified whether the robot made an object detection error or not in 90.3 % of cases (Section 6.6.3), this shows that all three conditions can support error detection by users. However, modality-specific advantages and limitations exist, as described below.

Regarding advantages and limitations of the visual modality, the persistent nature of the visual representation makes it easier to keep track, compared to only speech, as reported in participant interviews. An advantage of the *Visualization* is that it is seen as faster by some, while disadvantages are that it requires being able to see well and its limitations for representing complex scenes. In the visualization condition, it is important to be able to handle all possible object configurations, as some visualization-specific errors were reported. The robot that was used in the experiment was the Pepper robot, which has a built-in tablet. One solution to transfer our findings to other (humanoid) robots that do not have an in-built screen is by developing a smartphone or tablet app that performs a similar function.

Regarding the auditory modality, the speech representation resulted in uncertainty in some cases.

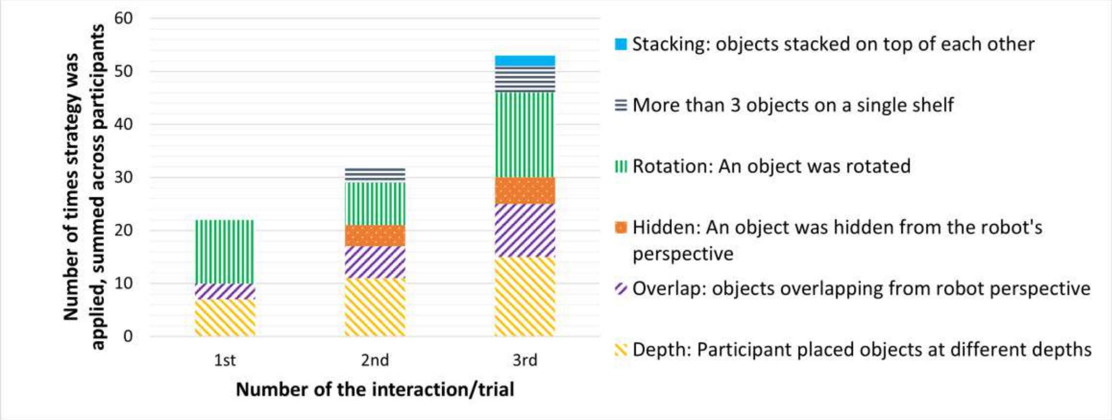


Figure 6.8: Number of times per interaction that participants applied object organization strategies with increased complexity, which made object detection more difficult. The strategies are summed up across all participants for the first, second and third time they interacted with the system. The frequency of complex strategies increased from the first to the third interaction (meaning that participants applied more strategies to “test the system" in the second trial as compared to the first, and applied even more strategies in the third and final trial). In other words, participants increasingly displayed testing behavior.

Once in the *Speech* condition, a participant wrote on the form that they were unsure whether an object was missing (this was counted as reporting an error). As described in Section 6.6.3, in two cases in the Speech condition, objects were not detected but the participant said there was no error (to the robot and on the form). In one case in the Speech condition, the appearance of unknown objects was not remarked on to the robot or on the survey form. This suggests some errors were missed or participants were unsure due to the representation in the *Speech* condition not being persistent. The response of the participant for whom three objects were missing arose due to the nature of the representation; asking whether a speech-only representation is correct has the risk of the participant confirming its correctness, when the representation only partially covers the object arrangement. Uttering one true statement can be interpreted as the representation being correct, even if there is missing information; if that statement is correct, the missing information does not make the representation incorrect. For example, when the robot only mentions one of the objects at the correct location, this can be seen as a correct statement even if other objects are not mentioned. However, this may change if participants are more motivated to ensure the full representation is correct. With visualization, on the other hand, there may be more of an expectation that the representation completely matches the scene when asked to confirm its correctness.

An exploratory data analysis indicated that participants with higher self-reported computer programming experience were more likely to prefer the visualization-only condition as compared to a combination of visualization and speech. This indicates that novice users may prefer both interaction modalities while advanced users who are familiar with the system might prefer visualization only. We assume that the reason for this is that participants with more computer programming experience are more familiar with tablets and data visualizations on screens. Our results suggest to communicate information using both visualization and speech when a user starts interacting with a device and to offer the possibility to disable one of these at a later point. In our interviews, the multimodal condition was preferred as it was a double confirmation, which enhances certainty. Starting with both modalities would also be preferable in terms of accessibility. Speech can be helpful in case a user does not see well, and visualization in case a user has difficulty hearing. This should be considered in the design of robot embodiments; not all robots intended for human interaction include a built-in screen. A built-in screen or a connection to an external tablet allows for supporting human-robot joint activities with a visual shared representation.

A general observation for experiments in which participants are asked to detect errors is that participants should be asked if there are errors in several different ways, as they may respond differently to a robot than what they report on a form or to a researcher. Moreover, participants may have difficulty noticing additional errors after they have noticed an error. As described in Section 6.6.3, in three cases multiple errors occurred but the participant only reported on one of the errors. This could indicate that some participants may have focused on one error and did not pay attention to other potential errors that could have occurred, once they encountered an error.

6.7.2 How do the Observed Failure Cases Align with HRI Understanding of Failure?

In this section, we describe how the errors that occurred in the user study fit into current HRI failure taxonomies and understanding of failure in HRI. When looking at the errors from a technical perspective in Section 6.6.3, the resulting failures (objects being missing as they were not detected) can be described as software failures according to, e.g., the taxonomy by [134]. However, when we take errors reported by participants and participant behavior into account, a more complex picture emerges of the failures that occurred. In line with the calls by [278, 309], a more human-centered perspective on errors comes to the fore. The interaction errors according to the study participants as described in Table 6.5, namely speech recognition errors, next behaviors triggered too soon, speech being recognized too late, long response times, confirming twice, and the robot not turning itself to the participant can be classified as social norm violations in the human-robot failure taxonomy [134], or as social errors according to the taxonomy of social errors in HRI [278]. However, adding social norm violations to a taxonomy is not enough to adequately describe all potential failure situations.

During our main study, we observed participant behavior that made the object detection more difficult (Section 6.6.5). This behavior was seen to increase as the interactions progressed. This may also have contributed to the increase in errors between the last and the first interaction (Section 6.6.3). Such behavior could be interpreted as the participants having a mental model of how the object detection works, forming a question in their mind (e.g., *"will the robot recognize the object if I hide it?"*), and then testing it in subsequent interactions. Participants also reported that they were testing the system (Section 6.6.5). Errors that were mentioned on the survey forms (Section 6.6.3) included that "hidden objects" were not detected. This shows that a subcategory of "human errors" exists that arise due to the human uncertainty about the potential effects of their actions (and neither due to deliberately acting in a way that sabotages the system, nor due to acting without any intention whatsoever). It also shows that errors exist that arise from a combination of technical and system design limitations, participant behavior, participant expectations and understanding of the system, and the configuration of objects in the environment. This makes it difficult to classify such failures according to the source of failure, which is the foundation of several failure taxonomies ([134, 38], see Section 6.3.2).

Human errors in the taxonomy by [134] include mistakes, slips, lapses, and deliberate violations, which are described as *"intentional illegitimate actions (e.g., directing the robot to run into a wall)"* [278, p.3]. We argue that the testing behavior is a deliberate action, but it is not a deliberate violation per se; it is not an action that has the sole aim to make the system fail but rather to find out if a particular situation will result in a technical failure. This is desirable user behavior in low-risk situations, as it will improve user understanding of the way the system works. If a failure occurs, it is the result of an intentional action that serves a productive purpose, e.g., improving the user's understanding of the system. This is related to the concept of trial-and-error behavior and its importance for user learning [37]. We argue that the opportunity for playful trial-and-error behavior, or productive failure, is something that should be considered in systems design and failure understanding in HRI. This concept is not a part of the current HRI failure taxonomies (e.g., [134, 278]). Usually, learning is not explicitly considered in HRI failure taxonomies. Learning mistakes are considered in taxonomies in other domains such as e-learning [218], where e-learning errors have been construed as potentially harmful by some authors, as it may lead to low computer self-efficacy or computer anxiety, while others construe error in a more positive light and see it as a part of the learning process. In low-risk HRI scenarios, we argue that errors in the context of trial-and-error behavior perform a positive role for user learning.

User learning to establish a more accurate mental model of the way the technology functions is especially relevant in the HRI context. The disconnect between people's estimates of robot perception and reasoning capabilities and the robot's actual capabilities has been referred to as *mental model discrepancy* [211], asymmetry in perception [100], and the perceptual belief problem in HRI [272]. As we observed that the way the robot's behavior was designed impacted participant understanding of the way the system works (Section 6.6.4), we argue that interaction design can scaffold trial-and-error behavior and the development of a correct mental model of system function. To facilitate the user in gradually developing a more accurate model of the robotic system, it is important that the robot gives appropriate behavioral cues. Similarly, [279] mention training and interaction design as potential failure mitigation strategies when a trust violation occurs. Interaction design can support user learning during interactions with a robotic system, where human interaction partners gradually develop a better understanding and expectations of the robotic system. For instance, this can be done through initial interactions in which object detection limitations are demonstrated, or by providing sensor information when the user detects an error.

6.8 Limitations of the Study

Our study does not address long-term effects, as the participants were asked to organize the objects only three times. Moreover, the realism of the scenario was limited. The shelf was initially empty, and the participants were asked to place a limited set of objects on the shelf. In a more realistic setup, the cupboard may already contain some objects. In such situations, it makes sense to display only newly added objects. Another limitation was that the robot was not capable of performing object manipulation. This may have impacted the participants' motivation to ensure that the robot correctly detected the objects. [134] discuss motivation as an important aspect for solving failures and mention the mitigation strategy of setting expectations regarding potential errors. Fewer errors may occur if participants have a higher level of motivation, if they are given more specific instructions for object organization, or given insight into the robot's perceptual processing. A different experimental setup, in which participants would be required to complete the task as fast as possible, may have led to a situation in which participants would not display trial-and-error behavior. The low task complexity and missing robot manipulation capabilities may have contributed to the observed floor effect for task load scores. Future work should investigate if higher task complexity will lead to more pronounced differences between conditions. As we ran the experiment with visitors to a museum of science and technology, there were regularly other people present in the room, e.g., friends and relatives. This could have influenced the participants' behavior, for which the video data may be analyzed (see e.g., [105]). Moreover, that the study took place in a museum of science and technology may also have contributed to the participants' motivation to test the system's capabilities and learn more about it, or this may have resulted in selection of participants with an interest in technology. Advantages of the study location were that the study participant pool was balanced in terms of gender and had a wide range in terms of participant age and programming experience. At the same time, the study was restricted to German speaking participants in Vienna, Austria. The number of participants was restricted due to practical constraints, which means the quantitative data gives limited information. However, the aim of the experiment was to gather qualitative data on interactions of participants with a system that produces non-wizarded errors.

6.9 Future Work

Future work for human-in-the-loop error detection should include studying the efficacy of alternate visualizations. For example, the user could be shown the camera stream with bounding boxes around detected objects when the user detects a failure (different from the one chosen in our study, as explained in Section 6.4.4). The human interaction partner may also be offered follow-up actions such as repositioning the robot, an option to manually correct learned object locations, or to view sensor data. The system can be extended by incorporating capabilities to detect and represent more complex object configurations (using all spatial prepositions mentioned by [111]), or object pose estimation to cover 3D. However, errors can still occur (e.g., due to occlusions). To deal with such limitations in technical systems, our work provides suggestions on how to support human-in-the-loop error detection. Future research needs to investigate how interaction design can support the user in developing a more accurate mental model of the robotic system.

6.10 Conclusion

In this work, we considered the problem of supporting human-in-the-loop error detection in an object organization task involving object detection functionality, a robotic system, and a human who organizes objects. In our study, we investigated if shared representations can support the participant's error detection task. We evaluated efficacy of different output modalities for the design of shared representations that the end user can inspect to detect errors of which the system is not aware. We used a functional object detection system in our study, to investigate the types of errors that are likely to occur in practice.

We found that visualization, speech and a combined condition all supported error detection. For the speech-only condition, we observed most cases of uncertainty of participants. However, this condition also sufficiently supported error detection. Most participants preferred the combined condition. An exploratory analysis of our results suggests that users with more programming expertise prefer visualization; but this needs to be investigated further. We recommend using both speech and visualization to decrease uncertainty, especially at the start of the interaction, and offer the option to switch off one of these modalities as the interaction proceeds. Moreover, using both modalities has advantages in terms of accessibility. Our study shows that a visualization of the robot's knowledge base, in addition to speech or by itself, is highly preferred in the context of a task that requires a human interaction partner of the robot's internal state. Thus, a built-in screen or external tablet with a visualization of the knowledge base supports human-robot joint activities.

Secondly, as participants were able to freely interact with our system, we were able to observe failure cases that play a role in user learning. Participants were observed to gradually make the task more difficult for the robot, to test and better understand the system. Failures arose due to participants' interpretation of and uncertainties regarding the robot's behavior, their motivation to find out whether their interpretation is correct, participant actions on the environment and the robot's subsequent perception of the environment. Such failure cases are not considered in current HRI failure taxonomies. When the user is sufficiently supported through the system's interaction design, these types of failures are likely to positively contribute to the user's understanding of the way the system works. Supporting trial-and-error behavior and designing robotic systems so that they help the user improve their mental model of the way the system works will help prevent failure in the long run. A related point we want to make, is that current failure taxonomies in HRI that categorize failures based on their source (e.g., a sensor or an action by a human) are likely to overlook how multiple sources can together lead to failure situations, or to miss certain failure cases altogether (e.g., the productive failure/trial-and-error behavior that we describe in this paper). We argue that failure in HRI should be understood as an interconnected phenomenon, where combinations of actions by different agents in the environment lead to failure.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

HF: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft; MH: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, review & editing; BS: Conceptualization, Data Curation, Methodology, Writing - original draft, review & editing PH: Software, Writing - review & editing; BI:

Conceptualization, Methodology, Funding acquisition, Supervision, Writing - original draft, review & editing; MV: Funding Acquisition, Supervision, Writing - review & editing.

Funding

This research was funded in part by FWF #ConnectingMinds grant CM 100-N (Caring Robots // Robotic Care) for research personnel and publication costs, and it established collaboration between authors and the museum where the study was conducted. The OPUS call in the Weave programme under the project number K/NCN/000142 (iChores) funded research personnel. The Polish National Agency for Academic Exchange under grant number K/PBI/000171 (NAWA) funded the establishment of the collaboration between the Polish and Austrian university involved, providing travel costs for the research exchange.

Acknowledgments

Thanks to Assistant Prof. Dr. Astrid Weiss for the invaluable feedback. Thanks to the Technisches Museum Wien for their support, especially Jürgen Öhlinger and Christopher Roither. We thank the participants in our study.

Tables

Topic of taxonomy	**Main categories**	Subcategory	Subtopic
UGV failures [38]	Physical failures	Effector Sensor Control system Power Communications	
	Human failures	Design failures Interaction failures	
HRI failure taxonomy [134]	Technical failures	Software failures	Design Failures Communication Failures Processing Failures
		Hardware failures	Effectors Sensors Power Control
	Interaction failures	Social norm violations	
		Human errors	Mistakes Slips Lapses Deliberate Violations
		Environment & Other agents	Group-Level Judgement

			Working Environment Organizational flaws
Faults in HRI/ robotics [258]	Interaction	Humans Agents and Robots Environment	
	Algorithms	Decision Making Behavior Execution Perception Localization and Mapping	
	Software	Decision Making Behavior Execution Perception Low Level	
	Hardware	Platform Sensors Manipulators Controller	
HRI failures [279]	Design failure		
	System failure	Hardware Software	
	Expectation failure	Commission failure (unexpected behavior by the robot) Omission failure(the robot fails to act in accordance with user expectations)	
	User failure	Intentional Unintentional	

Table 6.1: Overview of different failure taxonomies in the HRI literature

Topic of taxonomy	**Categories**
Social errors in HRI [278]	Breach in empathetic and emotional reactions Insufficient social skills Misunderstanding the user Insufficient communicative functions Breach in collaboration and prosociality
Output-oriented failure typology for HRI [309]	Logic failures (involve that there is output that conforms with expectation, but the output is wrong) Semantic failures (involve that the output format does not conform with expectation) Syntax failures (involve that there is no output)

Table 6.2: Overview of the taxonomy of social errors in HRI by [278] and typology by [309]

1. How was it for you?
2. Which version did you prefer?
    - Version A (the robot speaks out the object positions)
    - Version B (the robot shows the object locations visually on the tablet)
    - Version C (the robot shows the object locations visually on the tablet and speaks out the object positions)
3. Why?
4. How did the robot learn the positions of the objects? Please tell more about how you understood the way the robot functioned. There are no right or wrong answers, please convey your impression.
5. Did you have a specific reason for the way you organized the objects?
6. Did you notice anything unexpected during the interaction with the robot?
7. Do you have any suggestions for improvement?
8. What do you think this experiment was about?
9. Was anything in the instructions unclear?
10. Is there anything else that you would like to communicate to the researchers?

Table 6.3: Interview questions after the participant interacted with the system in all three conditions

Codes for analyzing responses to the questions “Which version did you prefer?" (Question 2) and “Why?" (Question 3) Both, because... ... it is a double confirmation ... then you both see and hear it ... it is the most clear ... speech-only takes more effort ... then you can better compare/memorize it ... it is more humanlike to have both hearing and seeing in the communication ... I did not like the speaking away ... then I can’t make mistakes ... I have a left-right weakness Tablet, because... ... it is faster ... I am used to tablets ... conditions with speech are more effort Speech, because... ... it is too much information to also see it on the tablet
Codes for analyzing responses to the question “How did the robot learn the positions?" (Question 4) Participant made reference to: • Scanning • The objects were programmed in before or shown before, or it had existing representation of the objects (e.g., samples) • Markers • Camera/photo/photographically • Seeing/with the eyes • Sensors • Program/programming • Shape of object • Determining the position of the shelves • Size of object • Ultrasound and infrared sensors • Done by human • With the text on the object
Codes for analyzing responses to the question “Did you have a specific reason for the way you organized the objects?" (Question 5) • No reason • Testing the system • Symmetrical, clear arrangement • Ordering objects by category (e.g., sticking, measuring, technical)

Table 6.4: Codes for analysis of interviews

Error	Number of times mentioned on a survey form
Unspecified	3
Object detection errors	
Object(s) not detected/recognized/shown	20
Hidden object not detected	4
Could not show that an object was placed behind another	3
Wrong object mentioned	2
Additional object shown	2
Object position error	1
Interaction errors	
Speech recognition error	4
Next behavior triggered too soon before participant spoke	4
Speech was recognized too late	2
Response time too long	1
It confirmed twice	1
Robot did not turn itself to me	1

Table 6.5: Errors as indicated by participants on the survey form

APPENDIX A

Clarification of contributions

In Appendix A, the contributions to the publications that are part of the dissertation are clarified.

A.1 Introduction (Chapter 1)

Publication: Frijns, H. A., Schürer, O. (2022) *Design as a Practice in Human-Robot Interaction Research*. Book chapter. In S. T. Köszegi, M. Vincze (Eds.), Trust in Robots (pp. 3–29). TU Wien Academic Press. https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_1

Clarification of the contribution: I conducted the literature search and wrote the chapter. My co-author Oliver Schürer gave me feedback as a thesis advisor.

This book chapter was not included as-is, but partially re-used to inform the current introduction chapter.

A.2 Chapter 2

Publication: Frijns, H.A., Schürer, O., Koeszegi, S.T. (2021) *Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI)*. Int J of Soc Robotics 15, 473–500, Issue date: March 2023.
https://doi.org/10.1007/s12369-021-00785-7

Clarification of the contribution: I conducted the literature search, conceptual analysis, developed the model, and wrote the paper. My co-authors Oliver Schürer and Sabine Köszegi gave me feedback as thesis advisors.

A.3 Chapter 3

Frijns, H.A., Vetter, R., Hirschmanner, M., Grabler, R., Vogel, L., Koeszegi, S.T. (2024) *Co-design of Robotic Technology with Care Home Residents and Care Workers*. In: PETRA'24 Conference Proceedings

Clarification of the contribution: I proposed the initial idea of the workshop series focusing on “technical modules" (software and hardware). Together with Ralf Vetter, and with input from the other co-authors, I further developed the concept of the workshop series. I wrote the publication and analyzed the data from the workshop series as reported in the paper. My co-authors gave feedback on my writing, proposed edits and approved of the final publication. I presented the study proposal to the research ethics committee with Ralf Vetter and Reinhard Grabler.

Ralf Vetter was core facilitator of the workshop series and was present at all 13 workshops. I was present at 11 of the workshops as a facilitator. Matthias Hirschmanner was present at 9 of the workshops as a facilitator. Reinhard Grabler was present at 6 of the workshops as a facilitator. Laura Vogel was present at 4 of the workshops as a facilitator. All facilitators and Muhammad Saleemi (as mentioned in the Acknowledgements section of the paper) were involved in revising the transcripts of the workshops that were generated using Whisper.

I took the lead in planning and running the workshop on Participatory Design with input from Ralf Vetter. Ralf Vetter took the lead in planning and running the workshop on Admission, with input from me. Ralf Vetter and I jointly prepared the workshop on Robots. Matthias Hirschmanner planned and ran the workshop on Computer Vision. Reinhard Grabler planned and ran the workshop on Conversational AI. Facilitators met regularly to give each other feedback on the development of workshops, and to discuss planning of workshops. Laura Vogel was involved in data analysis of the transcripts of the Participatory Design workshops. Sabine Köszegi obtained funding for the project Caring Robots // Robotic Care, writing the proposal for the original Caring Robots project with other authors, and gave feedback in the role of thesis advisor.

A.4 Chapter 4

Frijns, H.A., Schmidbauer, C. (2021) *Design Guidelines for Collaborative Industrial Robot User Interfaces*. In: Ardito, C., et al. Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12934. Springer, Cham. https://doi.org/10.1007/978-3-030-85613-7_28

Clarification of the contribution: I shaped the research process, conducted the literature search, did the evaluation study and conducted the interviews, and wrote the paper. My co-author Christina Schmidbauer, who is an expert in cobot systems for manufacturing, was involved in the affinity diagramming exercise, evaluated existing cobot systems with me in the modified heuristic evaluation, and was involved in the writing of the paper in a minor role. We also discussed the plan for the research process together on several occasions.

A.5 Chapter 5

Frijns, H.A., Stoeva, D., Gelautz, M., Schürer, O. (2024) *Programming Robot Animation Through Human Body Movement*. In: ARSO 2024 Conference Proceedings

Clarification of the contribution:

I conducted the literature search for related work, did the conceptual work on the design space, and wrote the paper. My co-authors gave me feedback on the system and study development, as well as on the paper throughout the writing process. I programmed the GUI and the system architecture for recording robot motion (which incorporates a human-humanoid imitation system), shaped the study design and conducted it with the programming experts. Together with Darja Stoeva and Oliver Schürer I conducted the study with the movement expert participants.

The imitation system that is incorporated in the system that was used for the paper, was a system that Darja Stoeva and I developed together. We closely collaborated for the kinematics module of this system, which is reported in *"Analytical Solution of Pepper's Inverse Kinematics for a Pose Matching Imitation System"* by Stoeva, D., Frijns, H.A., Gelautz M. and Schürer, O (see Sec. 1.8). Besides the kinematics module, the imitation system also contains a Websocket implementation for communicating between Python and C#, sending commands to the Pepper robot, and scaling and mapping functions to translate detected human joint positions to joint positions in the Pepper robot's workspace. Darja Stoeva and I collaborated on developing the mathematical mapping from human to robot. Darja Stoeva developed the concept of the imitation system and evaluation of its accuracy. She is the main author of the code for sensing human joint positions with Kinect, the Websocket implementation, and worked most on correcting and evaluating the mapping module and the kinematics module.

A.5.1 Copyright statement

"In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of TU Wien's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html *to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation."*

A.6 Chapter 6

Publication: Frijns, H.A., Hirschmanner, M., Sienkiewicz, B., Hönig, P., Indurkhyam B. and Vincze, M. (2024) *Human-In-The-Loop Error Detection in an Object Organization Task with a Social Robot*. Frontiers in Robotics and AI, Sec. Human-Robot Interaction, Volume 11. https://doi.org/10.3389/frobt.2024.1356827

Contributions statement according to the CRediT taxonomy (https://credit.niso.org/) that is integrated into the publication:

HF: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft; MH: Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, review & editing; BS: Conceptualization, Data Curation, Methodology, Writing - original draft, review & editing PH: Software, Writing - review & editing; BI: Conceptualization, Methodology, Funding acquisition, Supervision, Writing - original draft, review & editing; MV: Funding Acquisition, Supervision, Writing - review & editing.

Bibliography

[1] George Adamides, Georgios Christou, Christos Katsanos, Michalis Xenos, and Thanasis Hadzilacos. Usability guidelines for the design of robot teleoperation: A taxonomy. *IEEE Transactions on Human-Machine Systems*, 45(2):256–262, 2015. DOI:10.1109/THMS.2014.2371048.

[2] George Adamides, Christos Katsanos, Yisrael Parmet, Georgios Christou, Michalis Xenos, Thanasis Hadzilacos, and Yael Edan. HRI usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer. *Applied Ergonomics*, 62:237–246, 2017. DOI:10.1016/j.apergo.2017.03.008.

[3] Tanja Aitamurto, Donal Holland, and Sofia Hussain. Three layers of openness in design: Examining the open paradigm in design research. In *International Conference on Engineering Design*, 2013.

[4] Gopika Ajaykumar, Maureen Steele, and Chien-Ming Huang. A survey on end-user robot programming. *ACM Computing Surveys*, 54(8):1–36, 2022. DOI:10.1145/3466819.

[5] Gopika Ajaykumar, Maia Stiber, and Chien-Ming Huang. Designing user-centric programming aids for kinesthetic teaching of collaborative robots. *Robotics and Autonomous Systems*, 145:14, 2021. DOI:10.1016/j.robot.2021.103845.

[6] Morana Alač. Social robots: Things or agents? *AI & SOCIETY*, 31(4):519–535, November 2016. DOI:10.1007/s00146-015-0631-6.

[7] Ruben Albers, Judith Dörrenbächer, Martin Weigel, Dirk Ruiken, Thomas Weisswange, Christian Goerick, and Marc Hassenzahl. Meaningful telerobots in informal care: A conceptual design case. In *Nordic Human-Computer Interaction Conference*, pages 1–11. ACM, 2022. DOI:10.1145/3546155.3546696.

[8] Mina Alibeigi, Sadegh Rabiee, and Majid Nili Ahmadabadi. Inverse kinematics based human mimicking system using skeletal tracking technology. *Journal of Intelligent & Robotic Systems*, 85(1):27–45, 2017. DOI:10.1007/s10846-016-0384-6.

[9] Ruben Alonso, Alessandro Bonini, Diego Reforgiato Recupero, and Lucio Davide Spano. Exploiting virtual reality and the robot operating system to remote-control a humanoid robot. *Multimedia Tools and Applications*, 81(11):15565–15592, 2022. DOI:10.1007/s11042-022-12021-z.

[10] Patricia Alves-Oliveira, Kai Mihata, Raida Karim, Elin A. Bjorling, and Maya Cakmak. FLEX-SDK: An open-source software development kit for creating social robots. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–10. ACM, 2022. DOI:10.1145/3526113.3545707.

[11] Patrícia Alves-Oliveira, Alaina Orr, Elin A. Björling, and Maya Cakmak. Connecting the dots of social robot design from interviews with robot creators. *Frontiers in Robotics and AI*, 9:1–15, 2022. DOI:10.3389/frobt.2022.720799.

[12] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–13. ACM Press, 2019. DOI:10.1145/3290605.3300233.

[13] Shir Amir, Yossi Gandelsman, Shai Bagon, and Tali Dekel. Deep ViT features as dense visual descriptors. *arXiv:2112.05814*, 2022. DOI:10.48550/arXiv.2112.05814.

[14] Victor Nikhil Antony, Sue Min Cho, and Chien-Ming Huang. Co-designing with older adults, for older adults: Robots to promote physical activity. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 506–515. ACM, 2023. DOI:10.1145/3568162.3576995.

[15] James Auger. Living with robots: A speculative design approach. *Journal of Human-Robot Interaction*, 3(1):20, 2014. DOI:10.5898/JHRI.3.1.Auger.

[16] Minja Axelsson, Raquel Oliveira, Mattia Racca, and Ville Kyrki. Social robot co-design canvases: A participatory design framework. *ACM Transactions on Human-Robot Interaction*, 11(1):1–39, 2022. DOI:10.1145/3472225.

[17] Kasia Bail, Diane Gibson, Prativa Acharya, Julie Blackburn, Vera Kaak, Maria Kozlovskaia, Murray Turner, and Bernice Redley. Using health information technology in residential aged care homes: An integrative review to identify service and quality outcomes. *International Journal of Medical Informatics*, 165:104824, 2022. DOI:10.1016/j.ijmedinf.2022.104824.

[18] Etienne Balit, Dominique Vaufreydaz, and Patrick Reignier. Integrating animation artists into the animation design of social robots an open-source robot animation software. In *Proceedings of HRI'16*, pages 417–418, 2016. DOI:10.1109/HRI.2016.7451784.

[19] Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008. DOI:10.1080/10447310802205776.

[20] Kim Baraka, Patrícia Alves-Oliveira, and Tiago Ribeiro. An extended framework for characterizing social robots. *arXiv:1907.09873 [cs]*, pages 1–44, 2019.

[21] Dean C. Barnlund. A Transactional Model of Communication. In *Language Behavior: A Book of Readings in Communication*, pages 43–61. Mouton, The Hague, 1970.

[22] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. Design. In *Human-Robot Interaction: An Introduction*, pages 41–68. Cambridge University Press, 2020.

[23] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. What is human-robot interaction? In *Human-Robot Interaction: An Introduction*. Cambridge University Press, 2020.

[24] Andrea Bauer, Dirk Wollherr, and Martin Buss. Human-Robot Collaboration: A Survey. *International Journal of Humanoid Robotics*, 5(1):47–66, 2008. DOI:10.1142/S0219843608001303.

[25] Suna Bensch, Aleksandar Jevtić, and Thomas Hellström. On Interaction Quality in Human-Robot Interaction. In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*, pages 182–189, Porto, Portugal, 2017. SCITEPRESS - Science and Technology Publications. DOI:10.5220/0006191601820189.

[26] Peter L Berger and Thomas Luckmann. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Penguin Books, 1966.

[27] Laura Bishop, Anouk van Maris, Sanja Dogramadzi, and Nancy Zook. Social robots: The influence of human and robot characteristics on acceptance. *Paladyn, Journal of Behavioral Robotics*, 10(1):346–358, October 2019. DOI:10.1515/pjbr-2019-0028.

[28] Alan F. Blackwell. HCI as an inter-discipline. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 503–516. ACM, 2015. DOI:10.1145/2702613.2732505.

[29] Ann Bladford. Semi-structured qualitative studies. In *The Encyclopedia of Human-Computer Interaction*. The Interaction Design Foundation, 2 edition, 2013.

[30] Mike Blow, Kerstin Dautenhahn, Andrew Appleby, Chrystopher Nehaniv, and David Lee. Perception of robot smiles and dimensions for human-robot interaction design. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 469–474. IEEE, 2006. DOI:10.1109/ROMAN.2006.314372.

[31] Andrea Botero, Kari-Hans Kommonen, and Sanna Marttila. Expanding design space: Design-in-use activities and strategies. In *Design and Complexity - DRS International Conference 2010*, 2010.

[32] Tone Bratteteig and Ina Wagner. Design decisions and the sharing of power in PD. In *Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium papers, and Keynote abstracts - Volume 2*, pages 29–32. ACM, 2014. DOI:10.1145/2662155.2662192.

[33] Flor A Bravo, Alejandra M González, and Enrique González. A Review of Intuitive Robot Programming Environments for Educational Purposes. In *2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC)*, pages 1–6, Cartagena, Colombia, 2017. DOI:10.1109/CCAC.2017.8276396.

[34] Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. In *Springer Handbook of Robotics*, pages 1935–1971. Springer, 2nd edition, 2016.

[35] Daniel J Brooks. *A Human-Centric Approach to Autonomous Robot Failures*. University of Massachusetts Lowell, 2017. Dissertation.

[36] Maxime Busy and Maxime Caniot. qiBullet, a Bullet-based simulator for the Pepper and NAO robots. *arXiv preprint arXiv:1909.00779*, 2019.

[37] M. Cakmak and L. Takayama. Teaching people how to teach robots: The effect of instructional materials and dialog design. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 431–438, 2014. DOI:10.1145/2559636.2559675.

[38] J. Carlson and R.R. Murphy. How UGVs physically fail in the field. *IEEE Transactions on Robotics*, 21(3):423–437, 2005. DOI:10.1109/TRO.2004.838027.

[39] Felix Carros, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Honolulu HI USA, April 2020. ACM. DOI:10.1145/3313831.3376402.

[40] Felix Carros, Isabel Schwaninger, Adrian Preussner, Dave Randall, Rainer Wieching, Geraldine Fitzpatrick, and Volker Wulf. Care Workers Making Use of Robots: Results of a Three-Month Study on Human-Robot Interaction within a Care Home. In *CHI Conference on Human Factors in Computing Systems*, pages 1–15, New Orleans LA USA, April 2022. ACM. DOI:10.1145/3491102.3517435.

[41] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 541–548. IEEE, 2015. DOI:10.1109/ROMAN.2015.7333656.

[42] Daniel Chandler. The Transmission Model of Communication, 1994. http://visual-memory.co.uk/daniel/Documents/short/trans.html.

[43] Jessie Y. Chen, Katelyn Procci, Michael Boyce, Julia Wright, Andre Garcia, and Michael Barnes. Situation Awareness-Based Agent Transparency. Technical report, Defense Technical Information Center, Fort Belvoir, VA, April 2014. DOI:10.21236/ADA600351.

[44] Hui-Qing Chong, Ah-Hwee Tan, and Gee-Wah Ng. Integrated cognitive architectures: a survey. *Artificial Intelligence Review*, 28(2):103–130, 2007. DOI:10.1007/s10462-009-9094-9.

[45] H. Chung, M. Iorga, J. Voas, and S. Lee. “Alexa, Can I Trust You?”. *Computer*, 50(9):100–104, 2017. DOI:10.1109/MC.2017.3571053.

[46] Herbert H. Clark. Common ground. In *Using Language*. Cambridge University Press, 1996.

[47] Herbert H. Clark. Joint actions. In *Using Language*. Cambridge University Press, 1996.

[48] Herbert H. Clark. Joint Activities. In *Using Language*. Cambridge University Press, 1996.

[49] Herbert H. Clark. Meaning and understanding. In *Using Language*. Cambridge University Press, 1996.

[50] Edward Clarkson and Ronald C. Arkin. Applying heuristic evaluation to Human-Robot Interaction systems, 2007. American Association for Artificial Intelligence.

[51] Aurélie Clodic, Elisabeth Pacherie, Rachid Alami, and Raja Chatila. Key elements for human-robot joint action. In Raul Hakli and Johanna Seibt, editors, *Sociality and Normativity for Robots*, pages 159–177. Springer International Publishing, 2017. DOI:10.1007/978-3-319-53133-5_8.

[52] Mark Coeckelbergh. You, robot: on the linguistic construction of artificial others. *AI & SOCIETY*, 26(1):61–69, 2011. DOI:10.1007/s00146-010-0289-z.

[53] Mark Coeckelbergh. How to describe and evaluate "deception" phenomena: recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn. *Ethics and Information Technology*, 20(2):71–85, 2018. DOI:10.1007/s10676-017-9441-5.

[54] Philip R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 25:487–512, 1991.

[55] Enrique Coronado, Shunki Itadera, and Ixchel G. Ramirez-Alpizar. Integrating virtual, mixed, and augmented reality to human–robot interaction applications using game engines: A brief review of accessible software tools and frameworks. *Applied Sciences*, 13(3):1292, 2023. DOI:10.3390/app13031292.

[56] Enrique Coronado, Fulvio Mastrogiovanni, Bipin Indurkhya, and Gentiane Venture. Visual programming environments for end-user development of intelligent and social robots, a systematic review. *Elsevier Journal of Computer Languages*, 58(100970), 2020. DOI:10.1016/j.cola.2020.100970.

[57] Robert T. Craig. Communication theory as a field. *Communication Theory*, 9(2):119–161, 1999. DOI:10.1111/j.1468-2885.1999.tb00355.x.

[58] Arianna Curioni, Gunther Knoblich, and Natalie Sebanz. Joint action in humans: A model for human-robot interactions. In Ambarish Goswami and Prahlad Vadakkepat, editors, *Humanoid Robotics: A Reference*, pages 1–19. Springer Netherlands, 2017. DOI:10.1007/978-94-007-7194-9_126-1.

[59] Rick Dale, Riccardo Fusaroli, Nicholas D. Duran, and Daniel C. Richardson. The Self-Organization of Human Interaction. In *The Psychology of Learning and Motivation*, pages 43–96. Elsevier Inc.: Academic Press, 2014. DOI:10.1016/B978-0-12-407187-2.00002-2.

[60] Devleena Das, Siddhartha Banerjee, and Sonia Chernova. Explainable AI for Robot Failures: Generating Explanations that Improve User Assistance in Fault Recovery. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 351–360. ACM, 2021. DOI:10.1145/3434073.3444657.

[61] Devleena Das and Sonia Chernova. Semantic-Based Explainable AI: Leveraging Semantic Scene Graphs and Pairwise Ranking to Explain Robot Failures. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3034–3041. IEEE, 2021. DOI:10.1109/IROS51168.2021.9635890.

[62] Kerstin Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007. DOI:10.1098/rstb.2006.2004.

[63] Randall S. Davies. Understanding technology literacy: A framework for evaluating educational technology integration. *TechTrends*, 55(5):45–52, 2011. DOI:10.1007/s11528-011-0527-3.

[64] Ewart J. de Visser, Marieke M. M. Peeters, Malte F. Jung, Spencer Kohn, Tyler H. Shaw, Richard Pak, and Mark A. Neerincx. Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, 2019. DOI:10.1007/s12369-019-00596-x.

[65] Eric Deng, Bilge Mutlu, and Maja J Mataric. Embodiment in socially interactive robots. *Foundations and Trends in Robotics*, 7(4):251–356, 2019. DOI:10.1561/2300000056.

[66] Eric C Deng, Bilge Mutlu, and Maja J Matarić. Formalizing the design space and product development cycle for socially interactive robots. In *Workshop on Social Robots in the Wild at the 2018 ACM Conference on Human-Robot Interaction (HRI)*, page 6, 2018.

[67] Alan Dix. Designing for appropriation. In *Proceedings of the 21st BCS HCI Group Conference*, volume 2, pages 27–30, 2007.

[68] Anna Dobrosovestnova, Isabel Schwaninger, and Astrid Weiss. With a little help of humans. an exploratory study of delivery robots stuck in snow. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1023–1029. IEEE, 2022. DOI:10.1109/RO-MAN53752.2022.9900588.

[69] Fethiye Irmak Doğan, Gaspar I. Melsión, and Iolanda Leite. Leveraging explainability for understanding object descriptions in ambiguous 3D environments. *Frontiers in Robotics and AI*, 9:937772, 2023. DOI:10.3389/frobt.2022.937772.

[70] Anca D. Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S. Srinivasa. Effects of robot motion on human-robot collaboration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, pages 51–58. ACM Press, 2015. DOI:10.1145/2696454.2696473.

[71] drag&bot. Industrieroboter wie ein smartphone bedienen, 2020. https://www.dragandbot.com/de/, Last visited on 2020-09-10.

[72] Jill L. Drury, Dan Hestand, Holly A. Yanco, and Jean Scholtz. Design guidelines for improved human-robot interaction. In *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, page 1540. ACM Press, 2004. DOI:10.1145/985921.986116.

[73] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. Multimodal interfaces: A survey of principles, models and frameworks. In Denis Lalanne and Jürg Kohlas, editors, *Human Machine Interaction*, volume 5440, pages 3–26. Springer Berlin Heidelberg, 2009. DOI:10.1007/978-3-642-00437-7_1.

[74] Judith Dörrenbächer, Marc Hassenzahl, Robin Neuhaus, and Ronda Ringfort-Felner. Towards designing meaningful relationships with robots. In *Meaningful Futures with Robots—Designing a New Coexistence*, pages 3–29. Chapman and Hall/CRC, 2022. DOI:10.1201/9781003287445-1.

[75] Judith Dörrenbächer, Diana Löffler, and Marc Hassenzahl. Becoming a robot - overcoming anthropomorphism with techno-mimesis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2020. DOI:10.1145/3313831.3376507.

[76] European Data Protection Supervisor (EDPS). THE EDPS VIDEO-SURVEILLANCE GUIDELINES, 2010. https://edps.europa.eu/sites/edp/files/publication/10-03-17_video-surveillance_guidelines_en.pdf.

[77] Shirine El Zaatari, Mohamed Marei, Weidong Li, and Zahid Usman. Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems*, 116:162–180, June 2019. DOI:10.1016/j.robot.2019.03.003.

[78] Tatiana Aloi Emmanouil and Tony Ro. Amodal completion of unconsciously presented objects. *Springer Psychon Bull Rev*, pages 1188–1194, 2014. DOI:10.3758/s13423-014-0590-9.

[79] Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995. DOI:10.1518/001872095779049543.

[80] Publications Office of the European Union EUR-Lex Access to European Union Law. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation), 2016.

[81] Daniel Fallman. The interaction design research triangle of design practice, design studies, and design exploration. *Design Issues*, 24(3):4–18, 2008. DOI:10.1162/desi.2008.24.3.4.

[82] Janet Feigenspan, Christian Kästner, Jörg Liebig, Sven Apel, and Stefan Hanenberg. Measuring programming experience. In *2012 20th IEEE International Conference on Program Comprehension (ICPC)*, pages 73–82, June 2012. DOI:10.1109/ICPC.2012.6240511.

[83] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132:138–161, 2019. DOI:10.1016/j.ijhcs.2019.07.009.

[84] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamo-Larrieux. Robots and transparency: The multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics Automation Magazine*, 26(2):71–78, 2019. DOI:10.1109/MRA.2019.2904644.

[85] Federica Ferraguti, Andrea Pertosa, Cristian Secchi, Cesare Fantuzzi, and Marcello Bonfè. A methodology for comparative analysis of collaborative robots for industry 4.0. In *2019 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1070–1075, 2019. DOI:10.23919/DATE.2019.8714830.

[86] Julia Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams, editors, *Social Robotics*, volume 7621, pages 199–208. Springer Berlin Heidelberg, 2012. DOI:10.1007/978-3-642-34103-8_20.

[87] Gerhard Fischer and Elisa Giaccardi. Meta-design: A framework for the future of end-user development. In Henry Lieberman, Fabio Paternò, and Volker Wulf, editors, *End User Development - Empowering People to Flexibly Employ Advanced Information and Communication Technology*, volume 9 of *Human-Computer Interaction Series*, pages 427–457. Springer, 2006. DOI:10.1007/1-4020-5386-X_19.

[88] Kerstin Fischer. When Transparent does not Mean Explainable. In *Proceedings of 'Explainable Robotic Systems', Workshop in conjunction with the HRI 2018 Conference*, page 3, Chicago, 2017.

[89] Sarah R. Fletcher, Teegan L. Johnson, and Jon Larreina. Putting People and Robots Together in Manufacturing: Are We Ready? In Maria Isabel Aldinhas Ferreira, João Silva Sequeira, Gurvinder Singh Virk, Mohammad Osman Tokhi, and Endre E. Kadar, editors, *Robotics and Well-Being*, volume 95, pages 135–147. Springer International Publishing, 2019. DOI:10.1007/978-3-030-12524-0_12.

[90] Kristina Flägel, Britta Galler, Jost Steinhäuser, and Katja Götz. Der National Aeronautics and Space Administration-Task Load Index (NASA-TLX) – ein Instrument zur Erfassung der Arbeitsbelastung in der hausärztlichen Sprechstunde: Bestimmung der psychometrischen Eigenschaften. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 147-148:90–96, 2019. DOI:10.1016/j.zefq.2019.10.003.

[91] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166, 2003. DOI:10.1016/S0921-8890(02)00372-X.

[92] Terrence Fong, Charles Thorpe, and Charles Baur. Robot as Partner: Vehicle Teleoperation with Collaborative Control. In Alan C. Schultz and Lynne E. Parker, editors, *Multi-Robot Systems: From Swarms to Intelligent Automata*, pages 195–202. Springer Netherlands, Dordrecht, 2002. DOI:10.1007/978-94-017-2376-3_21.

[93] Jodi Forlizzi. How robotic products become social products: An ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 129–136, 2007.

[94] Jodi Forlizzi. The Product Ecology: Understanding Social Product Use and Supporting Design Culture. *International Journal of Design*, 2(1):11–20, 2008.

[95] Franka Emika GmbH. Franka Emika Panda, 2020. https://www.franka.de/technology.

[96] FraPorta. Pepper teleoperation using OpenPose, 2022. https://github.com/FraPorta/pepper_openpose_teleoperation.

[97] Christopher Frauenberger. Entanglement HCI the next wave? *ACM Transactions on Computer-Human Interaction*, 27(1):1–27, 2019. DOI:10.1145/3364998.

[98] Helena Anna Frijns, Anna Dobrosovestnova, Carina Brauneis, Reinhard Grabler, Matthias Hirschmanner, Darja Stoeva, Ralf Vetter, and Laura Vogel. Transdisciplinary and Participatory Research for Robotic Care Technology - Mapping Challenges and Perspectives. *The First Workshop on Care Robots for Older Adults (CROA), RO-MAN 2022*, page 3, 2022.

[99] Helena Anna Frijns and Christina Schmidbauer. Design guidelines for collaborative industrial robot user interfaces. In Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen, editors, *Human-Computer Interaction – INTERACT 2021*, volume 12934, pages 407–427. Springer International Publishing, 2021.

[100] Helena Anna Frijns, Oliver Schürer, and Sabine Theresia Koeszegi. Communication models in human–robot interaction: An asymmetric MODel of ALterity in human–robot interaction (AMODAL-HRI). *International Journal of Social Robotics*, page 28, 2021. DOI:10.1007/s12369-021-00785-7.

[101] Christian Fuchs. *Social Media, a critical introduction*. SAGE Publications, 2 edition, 2017.

[102] Norina Gasteiger, Ho Seok Ahn, Christopher Lee, Jongyoon Lim, Bruce A. MacDonald, Geon Ha Kim, and Elizabeth Broadbent. Participatory design, development, and testing of assistive health robots with older adults: An international four-year project. *ACM Transactions on Human-Robot Interaction*, 11(4):1–19, 2022. DOI:10.1145/3533726.

[103] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2918–2928, June 2021.

[104] Sarah Gillet, Marynel Vázquez, Sean Andrist, Iolanda Leite, and Sarah Sebo. Interaction-shaping robotics: Robots that influence interactions between other agents. *ACM Transactions on Human-Robot Interaction*, 13(1):1–23, 2024. DOI:10.1145/3643803.

[105] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Frontiers in Psychology*, 6, July 2015. DOI:10.3389/fpsyg.2015.00931.

[106] Franka Emika GmbH. User Handbook Panda, 2018.

[107] Erving Goffman. Primary frameworks. In *Frame Analysis - An Essay on the Organization of Experience*, pages 21–39. Northeastern University Press, 1974.

[108] Elizabeth Goodman, Erik Stolterman, and Ron Wakkary. Understanding interaction design practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1061–1070. ACM, 2011. DOI:10.1145/1978942.1979100.

[109] Michael A. Goodrich and Alan C. Schultz. Human-Robot Interaction: A survey. *Foundations and Trends® in Human-Computer Interaction*, 1(3):203–275, 2007. DOI:10.1561/1100000005.

[110] Jesse Gray, Guy Hoffman, Sigurdur Orn Adalgeirsson, Matt Berlin, and Cynthia Breazeal. Expressive, interactive robots: Tools, techniques, and insights based on collaborations. In *HRI 2010 Workshop: What do collaborations with the arts have to say about HRI*, page 8, 2010.

[111] Sergio Guadarrama, Lorenzo Riano, Dave Golland, Daniel Gouhring, Yangqing Jia, Dan Klein, Pieter Abbeel, and Trevor Darrell. Grounding spatial relations for human-robot interaction. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1640–1647. IEEE, 2013. DOI:10.1109/IROS.2013.6696569.

[112] David J Gunkel. Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century. *communication +1*, 1(1):26, 2012. DOI:10.7275/R5QJ7F7R.

[113] Andrea L Guzman and Seth C Lewis. Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, 22(1):70–86, 2020. DOI:10.1177/1461444819858691.

[114] Kasper Hald, Katharina Weitz, Elisabeth André, and Matthias Rehm. “An error occurred!” - trust repair with virtual robot using levels of mistake explanation. In *Proceedings of the 9th International Conference on Human-Agent Interaction*, pages 218–226. ACM, 2021. DOI:10.1145/3472307.3484170.

[115] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G. Pipe, and Kerstin Eder. Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 493–500, 2016. DOI:10.1109/ROMAN.2016.7745163.

[116] Zhao Han and Holly Yanco. Communicating missing causal information to explain a robot’s past behavior. *ACM Transactions on Human-Robot Interaction*, 12(1):1–45, 2023. DOI:10.1145/3568024.

[117] John W. Hansen. To change perceptions of technology programs. *Journal of Technology Studies*, 29(2):116–119, 2003. DOI:10.21061/jots.v29i2.a.10.

[118] Sandra G Hart. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*, 50(9):904–908, 2006.

[119] Marc Hassenzahl, Jan Borchers, Susanne Boll, Astrid M. Rosenthal v.d. Pütten, and Volker Wulf. Otherware: How to best interact with autonomous systems. *Interactions*, pages 54–57, 2021. DOI:10.1145/3436942.

[120] Frank Hegel. A modular interface design to indicate a robot's social capabilities. In *ACHI 2013: The Sixth International Conference on Advances in Computer-Human Interactions*, pages 426–432, 2013.

[121] Frank Hegel, Sebastian Gieselmann, Annika Peters, Patrick Holthaus, and Britta Wrede. Towards a typology of meaningful signals and cues in social robotics. In *2011 RO-MAN*, pages 72–78, Atlanta, GA, USA, July 2011. IEEE. DOI:10.1109/ROMAN.2011.6005246.

[122] Thomas Hellström and Suna Bensch. Understandable robots - What, Why, and How. *Paladyn, Journal of Behavioral Robotics*, 9(1):110–123, July 2018. DOI:10.1515/pjbr-2018-0009.

[123] Laura M Hiatt, Cody Narber, Esube Bekele, Sangeet S Khemlani, and J Gregory Trafton. Human modeling for human–robot collaboration. *The International Journal of Robotics Research*, 36(5-7):580–596, 2017. DOI:10.1177/0278364917690592.

[124] High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics guidelines for trustworthy AI, 2019.

[125] Matthias Hirschmanner, Stephanie Gross, Setareh Zafari, Brigitte Krenn, Friedrich Neubarth, and Markus Vincze. Investigating transparency methods in a robot word-learning system and their effects on human teaching behaviors. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 175–182, August 2021. DOI:10.1109/RO-MAN50785.2021.9515518.

[126] Matthias Hirschmanner, Christiana Tsiourti, Timothy Patten, and Markus Vincze. Virtual Reality Teleoperation of a Humanoid Robot Using Markerless Human Upper Body Pose Imitation. In *Humanoids 2019*, pages 259–265, October 2019. DOI:10.1109/Humanoids43949.2019.9035064.

[127] Guy Hoffman. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3):209–218, 2019. DOI:10.1109/THMS.2019.2904558.

[128] Guy Hoffman and Cynthia Breazeal. Collaboration in Human-Robot Teams. In *AIAA 1st Intelligent Systems Technical Conference*, Chicago, Illinois, September 2004. American Institute of Aeronautics and Astronautics. DOI:10.2514/6.2004-6434.

[129] Guy Hoffman and Wendy Ju. Designing robots with movement in mind. *Journal of Human-Robot Interaction*, 3(1):89, 2014. DOI:10.5898/JHRI.3.1.Hoffman.

[130] Guy Hoffman, Rony Kubat, and Cynthia Breazeal. A hybrid control system for puppeteering a live robotic stage actor. In *RO-MAN 2008*, pages 354–359, 2008. DOI:10.1109/ROMAN.2008.4600691.

[131] Guy Hoffman and Xuan Zhao. A Primer for Conducting Experiments in Human–Robot Interaction. *ACM Transactions on Human-Robot Interaction*, 10(1):1–31, 2020. DOI:10.1145/3412374.

[132] Laura Hoffmann, Nikolai Bock, and Astrid M. Rosenthal v.d. Pütten. The peculiarities of robot embodiment (EmCorp-scale): Development, validation and initial test of the embodiment and corporeality of artificial agents scale. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 370–378. ACM, 2018. DOI:10.1145/3171221.3171242.

[133] Lars Erik Holmquist and Jodi Forlizzi. Introduction to journal of human-robot interaction special issue on design. *Journal of Human-Robot Interaction*, 3(1):3, 2014. DOI:10.5898/JHRI.3.1.Holmquist.

[134] Shanee Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in Psychology*, 9:861, 2018. DOI:10.3389/fpsyg.2018.00861.

[135] Long-Jing Hsu, Janice K Bays, Katherine M. Tsui, and Selma Sabanovic. Co-designing social robots with people living with dementia: Fostering identity, connectedness, security, and autonomy. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pages 2672–2688. ACM, 2023. DOI:10.1145/3563657.3595987.

[136] Susan Hurley. The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences*, 31(1):1–22, February 2008. DOI:10.1017/S0140525X07003123.

[137] Kristina Höök and Jonas Löwgren. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction*, 19(3):1–18, 2012. DOI:10.1145/2362364.2362371.

[138] Don Ihde. *Technology and the Lifeworld, From Garden To Earth*. Indiana University Press, 1990.

[139] Don Ihde. *Postphenomenology and Technoscience, the Peking University lectures*. Suny Press, Albany, 2009.

[140] International Organization for Standardization (ISO). ISO 8373:2021(en), robotics — vocabulary, 2021.

[141] Tudor B. Ionescu and Sebastian Schlund. A participatory programming model for democratizing cobot technology in public and industrial fablabs. *Procedia CIRP*, 81:93–98, 2019. DOI:10.1016/j.procir.2019.03.017.

[142] ISO. ISO 9241-210 ergonomics of human–system interaction — part 210: Human-centred design for interactive systems, 2010.

[143] ISO. ISO/TS 15066:2016(en) Robots and robotic devices — Collaborative robots, 2016.

[144] Ryan Blake Jackson and Tom Williams. A theory of social agency for human-robot interaction. *Frontiers in Robotics and AI*, 8:687726, 2021. DOI:10.3389/frobt.2021.687726.

[145] Jerry A. Jacobs and Scott Frickel. Interdisciplinarity: A critical assessment. *Annual Review of Sociology*, 35(1):43–65, 2009. DOI:10.1146/annurev-soc-070308-115954.

[146] Glenn Jocher et al. ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference, 2022. DOI:10.5281/zenodo.6222936.

[147] Elizabeth Jochum and Jeroen Derks. Tonight we improvise!: Real-time tracking for human-robot improvisational dance. In *MOCO '19: 6th International Conference on Movement and Computing*, pages 1–11. ACM, 2019. DOI:10.1145/3347122.3347129.

[148] Ulla Johansson-Sköldberg, Jill Woodilla, and Mehves Çetinkaya. Design thinking: Past, present and possible futures. *Creativity and Innovation Management*, 22(2):121–146, 2013. DOI:10.1111/caim.12023.

[149] Matthew Johnson, Jeffrey M. Bradshaw, Paul J. Feltovich, Catholijn M. Jonker, M. Birna Van Riemsdijk, and Maarten Sierhuis. Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, 3(1):43, March 2014. DOI:10.5898/JHRI.3.1.Johnson.

[150] Malte Jung and Pamela Hinds. Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 7(1):5, 2018.

[151] ketchart. Pepper-robot-controlled-by-kinect-in-ubuntu, 2020. DOI:10.5281/zenodo.3828158.

[152] S. Kiesler. Fostering common ground in human-robot interaction. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 729–734, 2005. DOI:10.1109/ROMAN.2005.1513866.

[153] Jaewoo Kim, Kyoung Soo Chun, and Dong-Soo Kwon. Gesture motion programming by applying robot motion hierarchy structure for the educational/entertainment robot engkey. In *2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 36–39, 2012. DOI:10.1109/ARSO.2012.6213395.

[154] D. Lawrence Kincaid. The convergence model of communication. *Papers of the East-West Communication Institute*, No. 18:52, 1979.

[155] H. Knight and R. Simmons. An intelligent design interface for dancers to teach robots. In *RO-MAN 2017)*, pages 1344–1350, 2017. DOI:10.1109/ROMAN.2017.8172479.

[156] Titanilla Komenda. SAMY - semi-automatische modifikation, 2020. https://www.fraunhofer.at/de/forschung/forschungsfelder/SAMY.html, Visited on 2020-09-10.

[157] Dimosthenis Kontogiorgos, Andre Pereira, Boran Sahindal, Sanne Van Waveren, and Joakim Gustafson. Behavioural responses to robot conversational failures. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 53–62. ACM, 2020. DOI:10.1145/3319502.3374782.

[158] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. Embodiment effects in interactions with failing robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14. ACM, 2020. DOI:10.1145/3313831.3376372.

[159] Iuliia Kotseruba and John K. Tsotsos. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53(1):17–94, 2020. DOI:10.1007/s10462-018-9646-y.

[160] Kamil Krot and Vitalii Kutia. Intuitive methods of industrial robot programming in advanced manufacturing systems. In Anna Burduk, Edward Chlebus, Tomasz Nowakowski, and Agnieszka Tubis, editors, *Intelligent Systems in Production Engineering and Maintenance*, volume 835, pages 205–214. Springer International Publishing, 2019. DOI:10.1007/978-3-319-97490-3_20.

[161] Geert-Jan M. Kruijff. Symbol grounding as social, situated construction of meaning in human-robot interaction. *KI - Künstliche Intelligenz*, 27(2):153–160, 2013. DOI:10.1007/s13218-013-0238-3.

[162] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, December 2013. DOI:10.1016/j.robot.2013.05.007.

[163] Nicole C. Krämer, Astrid von der Pütten, and Sabrina Eimler. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In Marielba Zacarias and José Valente de Oliveira, editors, *Human-Computer Interaction: The Agency Perspective*, volume 396, pages 215–240. Springer Berlin Heidelberg, 2012. DOI:10.1007/978-3-642-25691-2_9.

[164] U. Kuckartz and S. Rädiker. Die typenbildende qualitative inhaltsanalyse. In *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*, pages 176–195. Beltz Verlagsgruppe, Preselect.media GmbH, 5 edition, 2022.

[165] KUKA AG. LBR iiwa, 2020. https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa.

[166] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish Ram Nambiappan, Kodur Krishna Chaitanya, Ashwin Ramesh Babu, Joanne Mathew,

and Fillia Makedon. A survey of robots in healthcare. *Technologies*, 9(1):8, 2021. DOI:10.3390/technologies9010008.

[167] Stephanie J. Lackey, Daniel J. Barber, and Sushunova G. Martinez. Recommended considerations for human-robot interaction communication requirements. In Masaaki Kurosu, editor, *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, volume 8511, pages 663–674. Springer International Publishing, 2014. DOI:10.1007/978-3-319-07230-2_63.

[168] Amy LaViers, Catie Cuan, Catherine Maguire, Karen Bradley, Kim Brooks Mata, Alexandra Nilles, Ilya Vidrin, Novoneel Chakraborty, Madison Heimerdinger, Umer Huzaifa, Reika McNish, Ishaan Pakrasi, and Alexander Zurawski. Choreographic and somatic approaches for the development of expressive robotic systems. *Arts*, 7(2):11, 2018. DOI:10.3390/arts7020011.

[169] Hee Rin Lee, Fei Sun, Tariq Iqbal, and Brenda Roberts. Reimagining robots for dementia: From robots for care-receivers/giver to robots for carepartners. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 475–484. ACM, 2023. DOI:10.1145/3568162.3578624.

[170] Hee Rin Lee, Selma Šabanović, Wan-Ling Chang, Shinichi Nagata, Jennifer Piatt, Casey Bennett, and David Hakken. Steps toward participatory design of social robots: Mutual learning with older adults with depression. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pages 244–253. ACM Press, 2017. DOI:10.1145/2909824.3020237.

[171] John D Lee and Katrina A See. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):50–80, 2004. DOI:10.1518/hfes.46.1.50_30392.

[172] Juerg Lehni and Jonathan Puckey. Paper.js. 2011. http://paperjs.org/.

[173] Luís Leite and Veronica Orvalho. Shape your body: control a virtual silhouette using body motion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pages 1913–1918. ACM, 2012. DOI:10.1145/2212776.2223728.

[174] Elizabeth A. Lemerise and William F. Arsenio. An Integrated Model of Emotion Processes and Cognition in Social Information Processing. *Child Development*, 71(1):107–118, January 2000. DOI:10.1111/1467-8624.00124.

[175] Tsung-Yi Lin et al. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, pages 740–755, 2014.

[176] Alfred R. Lindesmith, Anselm L. Strauss, and Norman K. Denzin. *Social Psychology*. SAGE Publications, 8 edition, 1999.

[177] Stephen W. Littlejohn and Karen A. Foss. *Theories of human communication*. Waveland Press, 10th ed edition, 2011.

[178] Changsong Liu and Joyce Y Chai. Learning to mediate perceptual differences in situated human-robot dialogue. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2288–2294, 2015.

[179] Tamara Lorenz, Astrid Weiss, and Sandra Hirche. Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care. *International Journal of Social Robotics*, 8(1):125–143, 2016. DOI:10.1007/s12369-015-0325-8.

[180] Maria Luce Lupetti, Cristina Zaga, and Nazli Cila. Designerly Ways of Knowing in HRI: Broadening the Scope of Design-oriented HRI Through the Concept of Intermediate-level Knowledge. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 389–398, Boulder CO USA, March 2021. ACM. DOI:10.1145/3434073.3444668.

[181] I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer. The bielefeld anthropomorphic robot head “flobi”. In *2010 IEEE International Conference on Robotics and Automation*, pages 3384–3391, 2010. DOI:10.1109/ROBOT.2010.5509173.

[182] Ali Ahmad Malik and Arne Bilberg. Developing a reference model for human–robot interaction. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 2019. DOI:10.1007/s12008-019-00591-6.

[183] Jeremy A. Marvel, Shelly Bagchi, Megan Zimmerman, and Brian Antonishek. Towards effective interface designs for collaborative HRI in manufacturing: Metrics and measures. *ACM Transactions on Human-Robot Interaction*, 9(4):1–55, 2020. DOI:10.1145/3385009.

[184] Pauline Maurice, Ludivine Allienne, Adrien Malaise, and Serena Ivaldi. Ethical and Social Considerations for the Introduction of Human-Centered Technologies at Work. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 131–138, Genova, Italy, 2018. IEEE. DOI:10.1109/ARSO.2018.8625830.

[185] MAXQDA, VERBI GmbH. MAXQDA | All-In-One Qualitative & Mixed Methods Data Analysis Tool, 2020.

[186] Philipp Mayring. *Qualitative Inhaltsanalyse*, pages 159–175. Konstanz: UVK Univ.-Verl. Konstanz, 1994.

[187] Helinä Melkas, Lea Hennala, Satu Pekkarinen, and Ville Kyrki. Impacts of robot implementation on care personnel and clients in elderly-care institutions. *Elsevier International Journal of Medical Informatics*, 134, 2020. DOI:10.1016/j.ijmedinf.2019.104041.

[188] Joseph E. Michaelis, Amanda Siebert-Evenstone, David Williamson Shaffer, and Bilge Mutlu. Collaborative or simply uncaged? understanding human-cobot interactions in automation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2020. DOI:10.1145/3313831.3376547.

[189] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, 4:21, 2017. DOI:10.3389/frobt.2017.00021.

[190] Nicole Mirnig, Yeow Kee Tan, Tai Wen Chang, Yuan Wei Chua, Tran Anh Dung, Haizhou Li, and Manfred Tscheligi. Screen feedback in human-robot interaction: How to enhance robot expressiveness. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 224–230, 2014. DOI:10.1109/ROMAN.2014.6926257.

[191] Nicole Mirnig, Astrid Weiss, and Manfred Tscheligi. A communication structure for human-robot itinerary requests. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 205–206, 2011. DOI:10.1145/1957656.1957733.

[192] Robin R. Murphy and Satoshi Tadokoro. User interfaces for human-robot interaction in field robotics. In Satoshi Tadokoro, editor, *Disaster Robotics*, volume 128, pages 507–528. Springer International Publishing, 2019. DOI:10.1007/978-3-030-05321-5_11.

[193] Bilge Mutlu, Allison Terrell, and Chien-Ming Huang. Coordination Mechanisms in Human-Robot Collaboration. In *Proceedings of the HRI 2013 Workshop on Collaborative Manipulation*, page 6, 2013.

[194] Shin'ichiro Nakaoka. Choreonoid: Extensible virtual robot environment built on an integrated GUI framework. In *2012 IEEE/SICE International Symposium on System Integration (SII)*, pages 79–85, 2012. DOI:10.1109/SII.2012.6427350.

[195] Anja Naumann, Jörn Hurtienne, Johann Habakuk Israel, Carsten Mohs, Martin Christof Kindsmüller, Herbert A. Meyer, and Steffi Hußlein. Intuitive use of user interfaces: Defining a vague concept. In Don Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, volume 4562 of *Lecture Notes in Computer Science*, pages 128–136, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. DOI:10.1007/978-3-540-73331-7_14.

[196] Birthe Nesset, David A. Robb, José Lopes, and Helen Hastie. Transparency in HRI: Trust and decision making in the face of robot errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 313–317. ACM, 2021. DOI:10.1145/3434074.3447183.

[197] Robin Neuhaus, Ronda Ringfort-Felner, Judith Dörrenbächer, and Marc Hassenzahl. *How to Design Robots with Superpowers*, pages 43–54. Chapman and Hall/CRC, 1 edition, 2022. DOI:10.1201/9781003287445-3.

[198] Jakob Nielsen. Enhancing the explanatory power of usability heuristics. In *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 152–158, 1994.

[199] Donald A. Norman. *The design of everyday things*. Basic Books, New York, New York, revised and expanded edition, 2013.

[200] Samuel Olatunji, Tal Oron-Gilad, Vardit Sarne-Fleischmann, and Yael Edan. User-centered feedback design in person-following robots for older adults. *Paladyn, Journal of Behavioral Robotics*, 11(1):86–103, 2020. DOI:10.1515/pjbr-2020-0007.

[201] OpenAI. OpenAI platform, GPT-3.5, 2023. https://platform.openai.com/docs/models/gpt-3-5.

[202] OpenAI. Whisper, 2023. https://github.com/openai/whisper.

[203] OpenCV. OpenCV: Detection of ArUco markers. 2023. https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html.

[204] Anastasia K. Ostrowski, Cynthia Breazeal, and Hae Won Park. Long-term co-design guidelines: Empowering older adults as co-designers of social robots. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1165–1172. IEEE, 2021. DOI:10.1109/RO-MAN50785.2021.9515559.

[205] Antti Oulasvirta and Kasper Hornbæk. Counterfactual thinking: What theories *Do* in design. *International Journal of Human–Computer Interaction*, 38(1):78–92, 2022. DOI:10.1080/10447318.2021.1925436.

[206] Adam A. Pack. Language research: Dolphins. In Jennifer Vonk and Todd Shackelford, editors, *Encyclopedia of Animal Cognition and Behavior*, pages 1–10. Springer International Publishing, 2018.

[207] Dong Huk Park, Lisa Anne Hendricks, Zeynep Akata, Anna Rohrbach, Bernt Schiele, Trevor Darrell, and Marcus Rohrbach. Multimodal Explanations: Justifying Decisions and Pointing to the Evidence. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8779–8788. IEEE, 2018. DOI:10.1109/CVPR.2018.00915.

[208] Dhaval Parmar, Joseph Isaac, Sabarish V. Babu, Nikeetha D'Souza, Alison E. Leonard, Sophie Jörg, Kara Gundersen, and Shaundra B. Daily. Programming moves: Design and evaluation of applying embodied interaction in virtual environments to enhance computational thinking in middle school students. In *2016 IEEE Virtual Reality (VR)*, pages 131–140, 2016. DOI:10.1109/VR.2016.7504696.

[209] Chris Paxton, Andrew Hundt, Felix Jonathan, Kelleher Guerin, and Gregory D. Hager. CoSTAR: Instructing collaborative robots with behavior trees and vision. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 564–571, 2017. DOI:10.1109/ICRA.2017.7989070.

[210] Josef Perktold, Skipper Seabold, Kevin Sheppard, ChadFulton, Kerby Shedden, jbrockmendel, j grana6, Peter Quackenbush, Vincent Arel-Bundock, Wes McKinney, Ian Langmore, Bart Baker, Ralf Gommers, yogabonito, s scherrer, Yauhen Zhurko, Matthew Brett, Enrico Giampieri, yl565, Jarrod Millman, Paul Hobson, Vincent, Pamphile Roy, Tom Augspurger, tvanzyl, alexbrc, Tyler Hartley, Fernando Perez, Yuji Tamiya, and Yaroslav Halchenko. statsmodels/statsmodels: Release 0.14.1, 2023. DOI:10.5281/ZENODO.593847.

[211] Leah Perlmutter, Eric Kernfeld, and Maya Cakmak. Situated Language Understanding with Human-like and Visualization-Based Transparency. In *Robotics: Science and Systems XII*. Robotics: Science and Systems Foundation, 2016. DOI:10.15607/RSS.2016.XII.040.

[212] Gabriele Peters. *A View-Based Approach to Three-Dimensional Object Perception*. Universität Bielefeld, 2001. Dissertation.

[213] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F. Malle. What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 105–113. ACM, 2018. DOI:10.1145/3171221.3171268.

[214] Martin J. Pickering and Simon Garrod. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347, 2013. DOI:10.1017/S0140525X12001495.

[215] Pilotfabrik TU Wien. Pilot Factory TU Vienna – Industry 4.0, 2021.

[216] David Porfirio, Evan Fisher, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. Bodystorming Human-Robot Interactions. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 479–491, New Orleans LA USA, October 2019. ACM. DOI:10.1145/3332165.3347957.

[217] David J. Porfirio, Laura Stegner, Maya Cakmak, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. Figaro: A tabletop authoring environment for human-robot interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15. ACM, 2021. DOI:10.1145/3411764.3446864.

[218] Jason Priem. Fail better: Toward a taxonomy of e-learning error. *Journal of Educational Computing Research*, 43(4):533–535, 2010. DOI:10.2190/EC.43.3.f.

[219] Astrid Rosenthal-Von Der Pütten and Nikolai Bock. Development and validation of the self-efficacy in human-robot-interaction scale (SE-HRI). *ACM Transactions on Human-Robot Interaction*, 7(3):1–30, 2018. DOI:10.1145/3139352.

[220] Daniela Quiñones and Cristian Rusu. How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, 53:89–122, 2017. DOI:10.1016/j.csi.2017.03.009.

[221] Natasha Randall, Selma Šabanović, and Wynnie Chang. Engaging older adults with depression as co-designers of assistive in-home robots. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 304–309. ACM, 2018. DOI:10.1145/3240925.3240946.

[222] Harish Ravichandar, Athanasios S. Polydoros, Sonia Chernova, and Aude Billard. Recent Advances in Robot Learning from Demonstration. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):297–330, 2020. DOI:10.1146/annurev-control-100819-063206.

[223] James T. Reason. *Human error*. Cambridge Univ. Press, 20. print edition, 2009.

[224] Johan Redström. RE:definitions of use. *Design Studies*, 29(4):410–423, 2008. DOI:10.1016/j.destud.2008.05.001.

[225] Johan Redström. *Making design theory*. Design thinking, design theory. The MIT Press, Cambridge, Massachusetts, 2017.

[226] Byron Reeves and Clifford Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.

[227] Tiago Ribeiro and Ana Paiva. The practice of animation in robotics. In Nicoletta Noceti, Alessandra Sciutti, and Francesco Rea, editors, *Modelling Human Motion*, pages 237–269. Springer International Publishing, 2020. DOI:10.1007/978-3-030-46732-6_12.

[228] Finn Rietz, Alexander Sutherland, Suna Bensch, Stefan Wermter, and Thomas Hellström. WoZ4u: An open-source wizard-of-oz interface for easy, efficient and robust HRI experiments. *Frontiers in Robotics and AI*, 8:668057, 2021. DOI:10.3389/frobt.2021.668057.

[229] J. Rios-Martinez, A. Spalanzani, and C. Laugier. From Proxemics Theory to Socially-Aware Navigation: A Survey. *International Journal of Social Robotics*, 7(2):137–153, April 2015. DOI:10.1007/s12369-014-0251-1.

[230] Judy Robertson and Maurits Kaptein, editors. *Modern Statistical Methods for HCI*. Human–Computer Interaction Series. Springer International Publishing, 2016. DOI:10.1007/978-3-319-26633-6.

[231] Robotiq. Products: Grippers, Camera and Force Torque Sensors, 2021. https://robotiq.com/products.

[232] Universal Robots. UR5/CB3 original instructions (en), 2019.

[233] Wendy A. Rogers, Travis Kadylak, and Megan A. Bayles. Maximizing the benefits of participatory design for human–robot interaction research with older adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 64(3):441–450, 2022. DOI:10.1177/00187208211037465.

[234] Jeremy Rose and Matthew Jones. The double dance of agency: A socio-theoretic account of how machines and humans interact. *Systems, Signs & Actions, An International Journal on Communication, Information Technology and Work*, 1(1):19–37, 2005.

[235] Pierre Rouanet, Jerome Bechu, and Pierre-Yves Oudeyer. A comparison of three interfaces using handheld devices to intuitively drive and show objects to a social robot: the impact of underlying metaphors. In *RO-MAN 2009*, pages 1066–1072. IEEE, 2009. DOI:10.1109/ROMAN.2009.5326260.

[236] William B Rouse and Nancy M Morris. On looking into the black box: Prospects and limits in the search for mental models. Technical report, Center for Man-Machine Systems Research, School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta GA 30332, 1985.

[237] Stuart J. Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Prentice Hall series in artificial intelligence. Prentice Hall, 1995.

[238] Elie Saad, Joost Broekens, and Mark A. Neerincx. An iterative interaction-design method for multi-modal robot communication. In *RO-MAN 2020*, pages 690–697, 2020. DOI:10.1109/RO-MAN47096.2020.9223529.

[239] M. Saerbeck and A. J. N. van Breemen. Design guidelines and tools for creating believable motion for personal robots. In *RO-MAN 2007*, pages 386–391, 2007. DOI:10.1109/ROMAN.2007.4415114.

[240] Eleanor Sandry. *Robots and communication*. Palgrave pivot. Palgrave Macmillan, 2015.

[241] Andrea Sanna, Fabrizio Lamberti, Gianluca Paravati, and Felipe Domingues Rocha. A kinect-based interface to animate virtual characters. *Journal on Multimodal User Interfaces*, 7(4):269–279, December 2013. DOI:10.1007/s12193-012-0113-9.

[242] Markus Schlosser. Agency. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2019 edition, 2019.

[243] Christina Schmidbauer, Titanilla Komenda, and Sebastian Schlund. Teaching cobots in learning factories – user and usability-driven implications. *Procedia Manufacturing*, 45:398–404, 2020. DOI:10.1016/j.promfg.2020.04.043.

[244] Eike Schneiders, EunJeong Cheon, Jesper Kjeldskov, Matthias Rehm, and Mikael B. Skov. Non-dyadic interaction: A literature review of 15 years of human-robot interaction conference publications. *ACM Transactions on Human-Robot Interaction*, 11(2):1–32, 2022. DOI:10.1145/3488242.

[245] Trenton Schulz, Jim Torresen, and Jo Herstad. Animation techniques in human-robot interaction user studies: A systematic literature review. *ACM Transactions on Human-Robot Interaction*, 8(2):1–22, 2019. DOI:10.1145/3317325.

[246] Svenja Y. Schött, Rifat Mehreen Amin, and Andreas Butz. A literature survey of how to convey transparency in co-located human–robot interaction. *Multimodal Technologies and Interaction*, 7(3):25, 2023. DOI:10.3390/mti7030025.

[247] Johanna Seibt. Classifying Forms and Modes of Co-Working in the Ontology of Asymmetric Social Interactions (OASIS). *Frontiers in Artificial Intelligence and Applications*, pages 133–146, 2018. DOI:10.3233/978-1-61499-931-7-133.

[248] Claude Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, first paperbound edition, tenth printing edition, 1964.

[249] E. Sibirtseva, D. Kontogiorgos, O. Nykvist, H. Karaoguz, I. Leite, J. Gustafson, and D. Kragic. A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 43–50, 2018. DOI:10.1109/ROMAN.2018.8525554.

[250] Bruno Siciliano and Oussama Khatib, editors. *Springer handbook of robotics*. Springer-Verlag Berlin Heidelberg, Berlin, 2008. DOI:10.1007/978-3-540-30301-5.

[251] David Sirkin and Wendy Ju. Using embodied design improvisation as a design research tool. In *International Conference on Human Behavior in Design*, page 7, 2014.

[252] Gillian Crampton Smith. What is interaction design? In *Designing Interactions*, pages vii–xix. The MIT Press, 2006.

[253] Softbank Robotics. Technical overview — aldebaran 2.5.11.14a documentation, 2017. http://doc.aldebaran.com/2-5/family/pepper_technical/index_pep.html, Last visited on 2020-08-07.

[254] SoftBank Robotics. Softbank robotics documentation, 2021. http://doc.aldebaran.com/2-5/index_dev_guide.html.

[255] SoftBank Robotics. Python SDK - overview — aldebaran 2.5.11.14a documentation. 2023. http://doc.aldebaran.com/2-5/dev/python/index.html.

[256] Katta Spiel. The Bodies of TEI – Investigating Norms and Assumptions in the Design of Embodied Interaction. In *TEI '21: Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–19, 2021. DOI:10.1145/3430524.3440651.

[257] Laura Stegner, Emmanuel Senft, and Bilge Mutlu. Situated participatory design: A method for in situ design of robotic interaction with older adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15. ACM, 2023. DOI:10.1145/3544548.3580893.

[258] Gerald Steinbauer. A survey about faults of robots used in RoboCup. In Xiaoping Chen, Peter Stone, Luis Enrique Sucar, and Tijn Van Der Zant, editors, *RoboCup 2012: Robot Soccer World Cup XVI*, volume 7500, pages 344–355. Springer Berlin Heidelberg, 2013. DOI:10.1007/978-3-642-39250-4_31.

[259] Franz Steinmetz, Annika Wollschläger, and Roman Weitschat. RAZER—a HRI for visual task-level programming and intuitive skill parameterization. *IEEE Robotics and Automation Letters*, 3(3):1362–1369, 2018. DOI:10.1109/LRA.2018.2798300.

[260] Darja Stoeva, Helena Anna Frijns, Oliver Schürer, and Margrit Gelautz. Analytical Solution of Pepper's Inverse Kinematics for a Pose Matching Imitation System. In *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, pages 167–174, August 2021. ISSN: 1944-9437.

[261] Erik Stolterman. The nature of design practice and implications for interaction design research. *International Journal of Design*, 2(1):55–65, 2008.

[262] Lucy Suchman. Human/machine reconsidered. *Cognitive Studies*, 5(1), 1998.

[263] Michael Suguitan and Guy Hoffman. Blossom: A handcrafted open-source robot. *ACM Transactions on Human-Robot Interaction*, 8(1):1–27, 2019. DOI:10.1145/3310356.

[264] JaYoung Sung, Rebecca E. Grinter, and Henrik I. Christensen. Domestic Robot Ecology: An Initial Framework to Unpack Long-Term Acceptance of Robots at Home. *International Journal of Social Robotics*, 2(4):417–429, December 2010. DOI:10.1007/s12369-010-0065-8.

[265] Rick Szostak, Claudio Gnoli, and María López-Huertas. *Interdisciplinary Knowledge Organization*. Springer International Publishing, 2016. DOI:10.1007/978-3-319-30148-8.

[266] Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater, and Florentin Wörgötter. Symbol emergence in cognitive developmental systems: A survey. *IEEE Transactions on Cognitive and Developmental Systems*, 11(4):494–516, 2019. DOI:10.1109/TCDS.2018.2867772.

[267] Thora Tenbrink, Evelyn Bergmann, Christoph Hertzberg, and Carsten Gondorf. Time will not help unskilled observers to understand a cluttered spatial scene. *Spatial Cognition & Computation*, 16(3):192–219, 2016. DOI:10.1080/13875868.2016.1143474.

[268] Thora Tenbrink, Kenny R. Coventry, and Elena Andonova. Spatial strategies in the description of complex configurations. *Discourse Processes*, 48(4):237–266, 2011. DOI:10.1080/0163853X.2010.549452.

[269] Moritz Tenorth and Michael Beetz. Representations for robot knowledge in the KnowRob framework. *Artificial Intelligence*, 247:151–169, 2017. DOI:10.1016/j.artint.2015.05.010.

[270] The European Parliament and the Council of the European Union. Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast). *Official Journal of the European Union*, 2006.

[271] The SciPy community. scipy.stats.friedmanchisquare — SciPy v1.12.0 manual, 2008-2024.

[272] Sam Thellman and Tom Ziemke. Do you see what I see? tracking the perceptual beliefs of robots. *iScience*, 23(10):101625, 2020. DOI:10.1016/j.isci.2020.101625.

[273] Andreas Theodorou, Robert H. Wortham, and Joanna J. Bryson. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science*, 29(3):230–241, 2017. DOI:10.1080/09540091.2017.1310182.

[274] A. L. Thomaz, M. Berlin, and C. Breazeal. An embodied computational model of social referencing. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 591–598, 2005. DOI:10.1109/ROMAN.2005.1513844.

[275] Andrea Thomaz, Guy Hoffman, and Maya Cakmak. Computational human-robot interaction. *Foundations and Trends in Robotics*, 4(2):104–223, 2016. DOI:10.1561/2300000049.

[276] Sofia Thunberg and Tom Ziemke. Social robots in care homes for older adults: Observations from participatory design workshops. In Haizhou Li, Shuzhi Sam Ge, Yan Wu, Agnieszka Wykowska, Hongsheng He, Xiaorui Liu, Dongyu Li, and Jairo Perez-Osorio, editors, *Social Robotics*, volume 13086, pages 475–486. Springer International Publishing, 2021. DOI:10.1007/978-3-030-90525-5_41.

[277] Leimin Tian, Pamela Carreno-Medrano, Aimee Allen, Shanti Sumartojo, Michael Mintrom, Enrique Coronado Zuniga, Gentiane Venture, Elizabeth Croft, and Dana Kulic. Redesigning Human-Robot Interaction in response to robot failures: a participatory design methodology. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–8. ACM, 2021. DOI:10.1145/3411763.3443440.

[278] Leimin Tian and Sharon Oviatt. A taxonomy of social errors in human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 10(2):1–32, 2021. DOI:10.1145/3439720.

[279] Suzanne Tolmeijer, Astrid Weiss, Marc Hanheide, Felix Lindner, Thomas M. Powers, Clare Dixon, and Myrthe L. Tielman. Taxonomy of trust-relevant failures and mitigation strategies. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 3–12. ACM, 2020. DOI:10.1145/3319502.3374793.

[280] J. Gregory Trafton, J. Malcolm McCurry, Kevin Zish, and Chelsea R. Frazier. The perception of agency. *ACM Transactions on Human-Robot Interaction*, 13(1):1–23, 2024. DOI:10.1145/3640011.

[281] Christiana Tsiourti, Astrid Weiss, Katarzyna Wac, and Markus Vincze. Designing emotionally expressive robots: A comparative study on the perception of communication modalities. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 213–222. ACM, 2017. DOI:10.1145/3125739.3125744.

[282] K. M. Tsui, K. Abu-Zahra, R. Casipe, J. M'Sadoques, and J. L. Drury. Developing heuristics for assistive robotics. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 193–194, March 2010. DOI:10.1109/HRI.2010.5453198.

[283] Universal Robots. UR5 collaborative robot arm | flexible and lightweight robot arm, 2019.

[284] Núria Vallès-Peris and Miquel Domènech. Caring in the in-between: a proposal to introduce responsible AI and robotics to healthcare. *AI & SOCIETY*, pages 1 – 11, 2021. DOI:10.1007/s00146-021-01330-w.

[285] Jeffrey Van Camp. My Jibo Is Dying and It's Breaking My Heart, August 2019. https://www.wired.com/story/jibo-is-dying-eulogy/.

[286] Sanne Van Waveren, Christian Pek, Jana Tumova, and Iolanda Leite. Correct me if I'm wrong: Using non-experts to repair reinforcement learning policies. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 493–501. IEEE, 2022. DOI:10.1109/HRI53351.2022.9889604.

[287] Aimee Van Wynsberghe and Shuhong Li. A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI). *Medicolegal and Bioethics*, Volume 9:11–21, 2019. DOI:10.2147/MB.S160348.

[288] Gentiane Venture and Dana Kulić. Robot expressive motions: A survey of generation and evaluation methods. *ACM Transactions on Human-Robot Interaction*, 8(4):1–17, 2019. DOI:10.1145/3344286.

[289] Peter-Paul Verbeek. Morality in design, design ethics and the morality of technological artifacts. In P.E. Vermaas, editor, *Philosophy and Design*, pages 91–103. Springer, 2008.

[290] Peter-Paul Verbeek. Beyond interaction: a short introduction to mediation theory. *Interactions*, 22(3):26–31, 2015. DOI:10.1145/2751314.

[291] Cordula Vesper, Ekaterina Abramova, Judith Bütepage, Francesca Ciardo, Benjamin Crossey, Alfred Effenberg, Dayana Hristova, April Karlinsky, Luke McEllin, Sari R. R. Nijssen, Laura Schmitz, and Basil Wahn. Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in Psychology*, 07, January 2017. DOI:10.3389/fpsyg.2016.02039.

[292] Valeria Villani, Fabio Pini, Francesco Leali, and Cristian Secchi. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, 55:248–266, 2018. DOI:10.1016/j.mechatronics.2018.02.009.

[293] Benjamin Walther-Franks and Rainer Malaka. An interaction approach to computer animation. *Entertainment Computing*, 5(4):271–283, December 2014. DOI:10.1016/j.entcom.2014.08.007.

[294] Chao Wang, Joerg Deigmoeller, Pengcheng An, and Julian Eggert. A user interface for sense-making of the reasoning process while interacting with robots. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7. ACM, 2023. DOI:10.1145/3544549.3585886.

[295] Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, and Vicente Ordonez. Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5309–5318. IEEE, 2019. DOI:10.1109/ICCV.2019.00541.

[296] David Weintrop, Afsoon Afzal, Jean Salac, Patrick Francis, Boyang Li, David C. Shepherd, and Diana Franklin. Evaluating CoBlox: A Comparative Study of Robotics Programming Environments for Adult Novices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal QC Canada, 2018. ACM. DOI:10.1145/3173574.3173940.

[297] Astrid Weiss, Regina Bernhaupt, Michael Lankes, and Manfred Tscheligi. The USUS evaluation framework for human-robot interaction. In *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction*, volume 4, page 8, 2009.

[298] Astrid Weiss, Daniela Wurhofer, Regina Bernhaupt, Martin Altmaninger, and Manfred Tscheligi. A methodological adaptation for heuristic evaluation of HRI. In *19th International Symposium in Robot and Human Interactive Communication*, pages 1–6. IEEE, 2010. DOI:10.1109/ROMAN.2010.5598735.

[299] Julika Welge and Marc Hassenzahl. Better than human: About the psychological superpowers of robots. In Arvin Agah, John-John Cabibihan, Ayanna M. Howard, Miguel A. Salichs, and

Hongsheng He, editors, *Social Robotics*, volume 9979, pages 993–1002. Springer International Publishing, 2016. DOI:10.1007/978-3-319-47437-3_97.

[300] A. William Evans, Matthew Marge, Ethan Stump, Garrett Warnell, Joseph Conroy, Douglas Summers-Stay, and David Baran. The future of human robot teams in the army: Factors affecting a model of human-system dialogue towards greater team collaboration. In Pamela Savage-Knepshield and Jessie Chen, editors, *Advances in Human Factors in Robots and Unmanned Systems*, volume 499, pages 197–209. Springer International Publishing, 2017. DOI:10.1007/978-3-319-41959-6_17.

[301] Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. What does the robot think? transparency as a fundamental design requirement for intelligent systems. In *Proceedings of the IJCAI Workshop on Ethics for Artificial Intelligence : International Joint Conference on Artificial Intelligence.*, page 7, 2016.

[302] Kim Wölfel, Jörg Müller, and Dominik Henrich. ToolBot: Robotically reproducing handicraft. In Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen, editors, *Human-Computer Interaction – INTERACT 2021*, volume 12934, pages 470–489. Springer International Publishing, 2021. DOI:10.1007/978-3-030-85613-7_32.

[303] Halina Sendera Mohd. Yakin and Andreas Totu. The semiotic perspectives of Peirce and Saussure: A brief comparative study. *Procedia - Social and Behavioral Sciences*, 155:4–8, 2014. DOI:10.1016/j.sbspro.2014.10.247.

[304] Haibin Yan, Marcelo H. Ang, and Aun Neow Poo. A Survey on Perception Methods for Human–Robot Interaction in Social Robots. *International Journal of Social Robotics*, 6(1):85–119, January 2014. DOI:10.1007/s12369-013-0199-6.

[305] H. A. Yanco and J. Drury. Classifying human-robot interaction: an updated taxonomy. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 2841–2846, October 2004. DOI:10.1109/ICSMC.2004.1400763.

[306] James E. Young, JaYoung Sung, Amy Voida, Ehud Sharlin, Takeo Igarashi, Henrik I. Christensen, and Rebecca E. Grinter. Evaluating Human-Robot Interaction: Focusing on the Holistic Interaction Experience. *International Journal of Social Robotics*, 3(1):53–67, January 2011. DOI:10.1007/s12369-010-0081-8.

[307] Setareh Zafari and Sabine T. Koeszegi. Machine agency in socio-technical systems: A typology of autonomous artificial agents. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 125–130, 2018. DOI:10.1109/ARSO.2018.8625765.

[308] Setareh Zafari and Sabine T. Koeszegi. Attitudes toward attributed agency: Role of perceived control. *International Journal of Social Robotics*, 2020. DOI:10.1007/s12369-020-00672-7.

[309] Xinyi Zhang, Sun Kyong Lee, Hoyoung Maeng, and Sowon Hahn. Effects of failure types on trust repairs in human–robot interactions. *International Journal of Social Robotics*, 15(9):1619–1635, 2023. DOI:10.1007/s12369-023-01059-0.

[310] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. An analysis and critique of research through design: towards a formalization of a research approach. In *DIS 2010*, pages 310–319, 2010. DOI:10.1145/1858171.1858228.

[311] Fernando Zuher and Roseli Romero. Recognition of human motions for imitation and control of a humanoid robot. In *2012 Brazilian Robotics Symposium and Latin American Robotics Symposium*, pages 190–195, 2012. DOI:10.1109/SBR-LARS.2012.38.

[312] Selma Šabanović. Robots in Society, Society in Robots: Mutual Shaping of Society and Technology as a Framework for Social Robot Design. *International Journal of Social Robotics*, 2(4):439–450, December 2010. DOI:10.1007/s12369-010-0066-7.

[313] Selma Šabanović, Marek P. Michalowski, and Linda R. Caporeal. Making friends: Building social robots through interdisciplinary collaboration. *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, page 7, 2007.